Crowther, Liam P., Luke, Sarah H., Arellano, Eduardo C., McCormack, Caitlin G., Ferreira, Vinina, Hillier, Jonathan, Heathcote, Richard, Kloen, Henk, Muñoz-Sáez, Andrés, Oliveira-Rebouças, Patricia and others (2024) *The Cool Farm Biodiversity metric: An evidence-based online tool to report and improve management of biodiversity at farm scale.* Ecological Indicators, 161 . ISSN 1470-160X.

The Cool Farm Biodiversity metric: An evidence-based online tool to report and improve management of biodiversity at farm scale

Liam P. Crowther [a], Sarah H. Luke [b,c], Eduardo C. Arellano [d,e], Caitlin G. McCormack [f], Vinina Ferreira [g], Jonathan Hillier [h], Richard Heathcote [i], Henk Kloen [j,k], Andrés Muñoz-Sáez [e,l], Patricia Oliveira-Rebouças [m], Fabiana Oliveira da Silva [n], Nadia Rojas-Arévalo [d,e], Natalia Zielonka [a], Lynn V. Dicks [a,b,*]

[a] *School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK*
[b] *Department of Zoology, University of Cambridge, CB2 3EJ, UK*
[c] *School of Biosciences, University of Nottingham, Sutton Bonington Campus, Nr Loughborough, LE12 5RD, UK*
[d] *Facultad de Agronomía y Sistemas Naturales & Instituto para el Desarrollo Sustentable, Pontificia Universidad Católica de Chile, Santiago 7830436, Chile*
[e] *Center of Applied Ecology and Sustainability (CAPES) Santiago 7820244, Chile*
[f] *3Keel, Fenlock Court, Long Hanborough, Oxfordshire, OX29 8LN, UK*
[g] *Universidade Federal do Vale do São Francisco, Petrolina, PE, 56300-990, UNIVASF, Brazil*
[h] *University of Edinburgh, The Royal (Dick) School of Veterinary Studies and The Roslin Institute Easter Bush Campus, Midlothian, EH25 9RG, UK*
[i] *Independent researcher*
[j] *Province of Gelderland*
[k] *CLM, Gutenbergweg 1, 4104 BA, Culemborg, Netherlands*
[l] *Facultad de Ciencias Agronómicas, Universidad de Chile, Santiago 8820808, Chile*
[m] *Departamento de Tecnologia e Ciências Sociais, Universidade do Estado da Bahia - Campus III (UNEB) Juazeiro, 48904-711, Brazil*
[n] *Universidade Federal de Sergipe, CEP49680-000 Nossa Senhora da Glória, Sergipe, Brazil/INCT-IN-TREE*

ARTICLE INFO

ABSTRACT

Halting biodiversity loss and achieving food security are both aims of the United Nations 2030 Agenda for Sustainable Development, but there is complex interplay between them. Agriculture drives biodiversity loss, but biodiversity provides benefits to agriculture. There is substantial potential to develop 'win-win' solutions for biodiversity and people within productive farmland, by boosting wildlife that can be supported, whilst maintaining yield and other services. To achieve this, farmers need to be able to assess the impacts of their management on biodiversity at farm scale. While suitable tools exist to drive improvement in biodiversity management, none incorporates evidence on the effectiveness of specific management practices. In this study we present the Cool Farm Biodiversity metric, which generates a farm-scale action-based biodiversity management assessment, scored using expert judgements and expert assessment of experimental evidence. The metric is designed to be biome-specific, so it responds to conservation aims, ecosystem processes and farming systems in particular biomes. To demonstrate that the metric is responsive to changes in farm management, we present an example of use on a large arable farm from the temperate forest biome.

1. Introduction

1.1. The need for biodiversity assessment tools

Biodiversity plays a key role in providing essential ecosystem services, contributing to clean water, carbon sequestration (Harrison et al., 2014; Tilman et al., 2006), soil maintenance, pest control, and pollination on which agricultural production depends (Brussaard et al., 2007; Karp et al., 2018; Kennedy et al., 2013; Tamburini et al., 2020). However, despite our awareness of its importance, biodiversity is declining globally, with land-use change and management intensification – largely for agricultural production – currently being the leading drivers of losses (e.g., Díaz et al., 2019; Butchart et al., 2010; Lambertini, 2020). Given its critical importance in service delivery, as well as

arguments related to inherent value, halting biodiversity loss is a key part of the United Nations 2030 Agenda for Sustainable Development. Specifically Sustainable Development Goal 15 states that we should: "Protect, restore and promote sustainable use of terrestrial ecosystems, […] reverse land degradation and halt biodiversity loss" (The United Nations General Assembly, 2015). In parallel to this, Sustainable Development Goal 2 aims to: "End hunger, achieve food security and improved nutrition, and promote sustainable agriculture" (The United Nations General Assembly, 2015). With growing global food consumption, and a current global agricultural system that already exceeds the Earth's biogeophysical limits (Gerten et al., 2020), it is critical that we find solutions to reconcile these competing demands, reversing biodiversity loss (Leclère et al., 2020) whilst also supporting long term food security (Searchinger et al., 2019).

Given the benefits that biodiversity can provide to agriculture, there is substantial potential to develop management strategies on farmland that allow win–win solutions for biodiversity and people – boosting both the wildlife that can be supported, and also yield and other services (e.g., Fischer et al., 2017; Clough et al., 2011; Cunningham et al., 2013). Options for more sustainable management can include strategies at both local and landscape scales. For example, these could include intensifying production in certain areas whilst leaving others as natural habitat at landscape scale ('land sparing', e.g., Phalan, 2018), maintaining natural habitat on less productive areas within a farm (e.g., Pywell et al., 2015), and intercropping (e.g., Li et al., 2020), or introducing flower resources or nest sites to support certain species at field scale (e.g., Garibaldi et al., 2014; Blaauw and Isaacs, 2014). Often, a combination of approaches, maintaining as much diversity at the landscape scale as possible, can be most successful (e.g., Kremen and Merenlender, 2018). Actions to conserve biodiversity on farmland also need to be tailored for local conditions and contexts, taking account of the potential for aspects of biodiversity, such as pests or invasive species, to impact negatively on agricultural production and native biodiversity (e.g., Herd-Hoare and Shackleton, 2020).

Whatever the strategy, given continuing declines in biodiversity on agricultural land (Rigal et al., 2023), greater uptake of biodiversity-friendly management strategies and substantial modifications to the way many farming systems currently operate are needed to reverse biodiversity loss. To meet this challenge, sustainability strategies employed by individual farms, corporate supply chains, and government-led policies all need to incorporate greater action to conserve biodiversity within productive farmland (e.g., Kremen and Merenlender, 2018). In this context, tools to drive improvement are at least as important as tools and models that predict actual outcomes for biodiversity.

A range of international and national government-led policies have sought to incentivise protection of biodiversity in farmland, from the 1980s onwards (e.g., agri-environment schemes in the European Union's Common Agricultural Policy; Batáry et al., 2015). The new Global Biodiversity Framework, adopted in December 2022 under the Convention on Biological Diversity, specifies '*a substantial increase of the application of biodiversity friendly practices*' as part of the target to manage agricultural areas sustainably by 2030 (Target 10; Convention on Biological Diversity 2022). Biodiversity has been incorporated into international industry guidelines for assessing the sustainability of farming practices (e.g., Sustainable Agriculture Initiative, 2014). There are now specific standards, measures, or reporting requirements related to biodiversity in a number of international certification schemes, such as Rainforest Alliance ("Rainforest Alliance," 2022), Global GAP ("Global GAP," 2022), the Roundtable on Sustainable Palm Oil ("Roundtable on Sustainable Palm Oil (RSPO)," 2022), and the Unilever Sustainable Agriculture Code (Smith, 2017).

A small number of software tools are available to measure and drive improvement in biodiversity performance of farms, and report against these standards, allowing farmers to assess the benefits of their current practices for biodiversity. These include the Gaia Biodiversity Yardstick

(Kloen, 2014), the European Biodiversity Performance Tool (EU LIFE Initiative, 2021), and the prototype New Zealand biodiversity assessment tool (MacLeod et al., 2018), which record and score habitat and farm management actions in terms of their biodiversity value. These tools are all designed for farmland in specific regions (Europe, or New Zealand) and are not intended to be used globally. The limited geographic scope means that they are not useful for organisations who wish to report or evaluate biodiversity impacts of farmed products across global supply chains. There are globally-applicable biodiversity tools that can be applied to farmed landscapes. NatCapMap ("NATCAP MAP," 2022) calculates natural capital values for a given landscape based on habitat areas and characteristics, while the GLOBIO 4 model ("GloBio," 2022; Schipper et al., 2020) quantifies human impacts on the intactness of biodiversity at a range of scales, using a modelling approach with a spatial resolution of 300 m. Both these tools can be used anywhere in the world, but they take no account of farm management actions. None of these existing tools goes beyond expert judgement to incorporate available scientific evidence for the effectiveness of different management options for biodiversity.

Including existing scientific evidence in conservation decision support tools, along with the capacity to update as new evidence emerges, is key to affecting real conservation change and using limited resources wisely (Sutherland et al., 2004; Dicks, Walsh, and Sutherland, 2014). Research has shown that many conservation decisions have historically been based on anecdote or advice from others, while evidence, especially peer-reviewed science, is rarely the first, most widely used or most valued source of knowledge among conservation managers (e.g., Cook, Hockings, and Carter, 2010; Young and Van Aarde, 2011; Kadykalo et al., 2021). Yet when data are collated and assessed systematically, the advice for best practices can often be different from what was assumed. Examples of this include 'beetle banks' in agriculture (Dicks et al., 2016) and 'bat gantries' over roads (Sutherland and Wordley, 2017). Initiatives such as Conservation Evidence ("Conservation Evidence," 2022) have made substantial progress in making scientific evidence freely available to stakeholders, but there is still much scope for this to be delivered in more focused formats, allowing stakeholders to see just the evidence that is relevant to their decisions (e.g., Shackelford et al., 2019).

1.2. Considerations for developing biodiversity assessment tools

Farming systems, and the natural ecosystems in which they are embedded, differ widely across the globe, according to environmental and biogeographic conditions. For example, an oil palm plantation in Indonesia, an irrigated mango farm in Brazil and an intensive apple orchard in Germany might all be described as 'perennial tree crop', or 'top fruit' systems, but the wild species they support, and the management actions that would be appropriate to conserve those species in the context of productive farming, are unlikely to be the same. If a software tool to assess biodiversity management on farms is to be global in scope, it must have the capacity to be adapted for different farming and biogeographic contexts. All its main elements, including its biodiversity objectives, the actions suggested, and the evidence and judgement used to derive scores, should be allowed to differ among contexts.

A tool for driving improvement in biodiversity management on farms needs to make assessments at a farm scale, rather than a wider landscape scale, because farms are the management units across which improvements can take place, and farmers, farm managers and advisors, are the actors capable of implementing management change (Sietz et al., 2022). Additionally, the majority of high-quality experimental evidence for the effectiveness of agricultural actions on biodiversity conservation comes from studies that test hypotheses at a farm- or field-scale (Dicks et al., 2014).

When designing a tool to drive improvements in biodiversity management at farm scale, there is a choice between *action-based* and *outcome-based* (or *results-based*) approaches to scoring. Following similar farm-scale biodiversity assessment tools cited above, we favour an

action-based approach where points are given for efforts made, rather than requiring measurements of biodiversity response. This is for two main reasons. Firstly, it is clear from the literature that the actual biodiversity found on a farm, in terms of number of species present and the community composition, strongly depends on landscape factors operating at scales larger than most farms (Gamez-Virues et al., 2015; Seibold et al., 2019; Tscharntke et al., 2012). The effectiveness of farm management actions for biodiversity has also been shown to depend on landscape context (e.g., Scheper et al., 2013). These landscape factors include, for example, heterogeneity of land uses or habitat types, proportion of semi-natural habitats, and edge densities, measured at scales of 1 km or more. They are largely (not entirely) outside an individual farmers' control or sphere of influence. It does not seem equitable to reward or penalise farmers for the broader context in which they find themselves, nor is it likely to drive improvement. Secondly, with the exception of areas of specific habitat types, measuring biodiversity itself (i.e., outcomes, such as numbers of species or individuals present) requires data inputs that farmers usually do not have, nor have the capacity to collect.

There are risks associated with a purely action-based approach, particularly if it is very prescriptive and does not allow managers to adjust or adapt to their contexts. For example, overly prescriptive agri-environment schemes led to declines, rather than increases, in the Danube clouded yellow butterfly, by incentivising synchronised mowing dates at large spatial scales (Konvicka et al., 2008). Wezel et al. (2018) found a majority of European mountain farmers (from 79 farmer interviews) would prefer results-based over action-based agri-environment measures, because they allow flexibility. These farmers did, however, perceive risks in implementing results-based measures, including a need for specialized biodiversity training. It may be possible to develop biodiversity monitoring protocols that allow farmers themselves, or lay people, to monitor biodiversity at farm scale without specialized knowledge (e.g., Tasser et al., 2019). To our knowledge, such approaches have only been tested in Europe.

To increase the chances of widespread biodiversity benefit, any decision support tool must be credible, user-friendly and fit for purpose in its intended decision-making context, providing results that are meaningful and actionable for the users. A study of factors that affected the uptake of decision support tools by farmers in the UK found that usability, cost-effectiveness, performance, relevance, and compatibility with compliance demands were key to determining how likely farmers were to use any particular tool (Rose et al., 2016). Factors such as cost-effectiveness and ease of use are even more critical in areas of the world where resources and access to technology are likely to be limited. Involving users – here, farmers, farm managers and agricultural advisers - in the design process is expected to increase the likelihood of uptake of decision support tools in agriculture (Rose et al., 2017).

Beyond the community of users for a specific decision support tool, there is a wider set of stakeholders - people who are affected by it or can influence its uptake or success. Evidence from a wide range of disciplines - including conservation and sustainability science - indicates that involving stakeholders effectively in design and decision-making on projects leads to greater benefits overall and higher probability of long-term success (Giakoumi et al., 2018; Jolibert and Wesselink, 2012; Kainer et al., 2009; C. J. MacLeod et al., 2022; Reed et al., 2018; Sterling et al., 2017). These benefits include providing a greater evidence base for decisions, giving greater public acceptance, higher chance of success and impact, and broader communication of initiatives (Haddaway et al., 2017). In particular, successful stakeholder engagement is often critical for ensuring that human well-being goals are being met, as well as purely environmental aims (Redpath et al., 2013). When developing a generic software tool to drive improvements in biodiversity management at farm scale, important stakeholders include supply chain managers, corporate sustainability experts, biodiversity conservation practitioners, land managers, farmers and researchers working on biodiversity in farmland.

To summarise, there is a clear need and demand from the agricultural industry for a tool to measure the performance of farms' actions for biodiversity conservation. Our goal is to meet this need with a farm-scale, location-specific tool, based on sound evidence, that can be used to drive improvements in practice. We aim to develop a tool that acts as a conduit for evidence not currently accessible to users, in a form that can incentivise good practice globally. We present the Cool Farm Biodiversity metric, a farm-scale scoring metric to measure improvement, built to the following specifications:

- Available globally, localised to diverse ecological and agricultural settings.
- Compiling data and reporting results at farm scale, rather than associated with individual products, or at larger landscape scale.
- Action- rather than outcome-based: users are scored for the actions they take to conserve biodiversity, with scoring that is responsive to evidence, without any attempt to directly measure biodiversity (with the exception of habitat areas).
- Easy to use for farmers: the majority of data input requires information that typical farmers already have, and the language is designed to be farmer-friendly.
- Developed in collaboration with users and stakeholders.

2. Methods

The Cool Farm Biodiversity metric is designed to provide a simple checklist of actions that can be adopted by farmers in different biomes. Users tick boxes according to actions taken on their farm and fill in details about the areas of different habitat types.

The output provides for each user a set of overall general biodiversity scores broken down by elements of their farm (see Section 2.1.5) and by species group, and a calculated proportional area for different broad habitat types. The actions that users can select and the species groups scored depend on the biome the farm is located in, which allows the metric to be responsive to differences in agricultural practices and conservation priorities across biomes.

2.1. Data sources and underlying techniques

2.1.1. The Conservation Evidence database

The Conservation Evidence database comprises plain-language summaries of > 8,400 individual scientific studies (as of 19 April 2023 see 'Conservation Evidence', 2022). These are compiled into > 3,600 actions, which are organised into synopses (Sutherland et al., 2019). Synopses are structured reviews in which experimental evidence for the effectiveness of interventions for conservation of a species group or habitat, or approaches to tackle a particular conservation issue, has been carefully assessed by a panel of experts in a two- or three-round modified Delphi process (Sutherland et al., 2019). For each intervention, a systematic manual literature search is used to collect documented experimental evidence from a range of sources, including peer-reviewed papers and grey literature. Each item of evidence (study) is described in a standardised summary format, with all summaries published online. The expert assessment places the intervention on three axes - effectiveness, certainty and harms - which allow interventions to be sorted into categories that are easy for practitioners to interpret, for example 'Likely to be beneficial'. Table 1 provides a full explanation of the categories, showing how they are derived from the axis scores, and how they are translated into evidence scores for specific actions in the Cool Farm Biodiversity metric.

2.1.2. Expert elicitation methods

The scoring that underlies the Cool Farm Biodiversity metric is created by eliciting two distinct responses from panels of experts. Firstly, 'expert judgement', where experts are asked to use their background and contextual knowledge to judge whether they think an action is likely to

Table 1
Correspondence between effectiveness categories, reported on 'Conservation Evidence' (2022), and evidence scores assigned to actions in the Cool Farm Biodiversity metric (CF-BM). Thresholds apply to median scores for effectiveness, certainty and harm, derived after a multi-round, iterative scoring process by an expert panel, following a modified Delphi method.

Category	Description	General criteria	Thresholds	CF-BM general score	CF-BM species group score
Beneficial	Effectiveness demonstrated by clear evidence. Expectation of harm small compared with benefits	High benefit score High certainty score Low harm score	Effectiveness: >60 % Certainty: >60 % Harm: <20 %	2	2
Likely to be beneficial	Effectiveness less well established than for those listed under 'beneficial' **OR** Clear evidence of moderate effectiveness	High benefit score Medium certainty score Low harm score **OR** Medium benefit score Medium or high certainty score Low harm score	Effectiveness: >60 % Certainty: 40–60 % Harm: <20 % **OR** Effectiveness: 40–60 % Certainty: ≥40 % Harm: <20 %	1	1
Trade-off between benefit and harms	Practitioners must weigh up beneficial and harmful effects according to circumstances and priorities	Medium or high benefit score Medium or high certainty score Medium or high harm score	Effectiveness: ≥40 % Certainty: ≥40 % Harm: ≥20 %	0	0
Unknown effectiveness	Currently insufficient data, or data of inadequate quality	Low certainty score	Effectiveness: Any Certainty: <40 % Harm: Any	0	0
Unlikely to be beneficial	Lack of effectiveness less well established than for those listed under 'likely to be ineffective or harmful'	Low benefit score Medium certainty score and/or some variation between experts	Effectiveness: <40 % Certainty: 40–60 % Harm: <20 %	0	0
Likely to be ineffective or harmful	Ineffectiveness or harmfulness demonstrated by clear evidence	Low benefit score High certainty score (regardless of harm) **OR** Low benefit score Medium or high certainty score Medium or high harm score	Effectiveness: <40 % Certainty: >60 % Harm: Any **OR** Effectiveness: <40 % Certainty: ≥40 % Harm: ≥20 %	REMOVE ACTION FROM LIST	−1

Adapted from Sutherland et al. (2019)
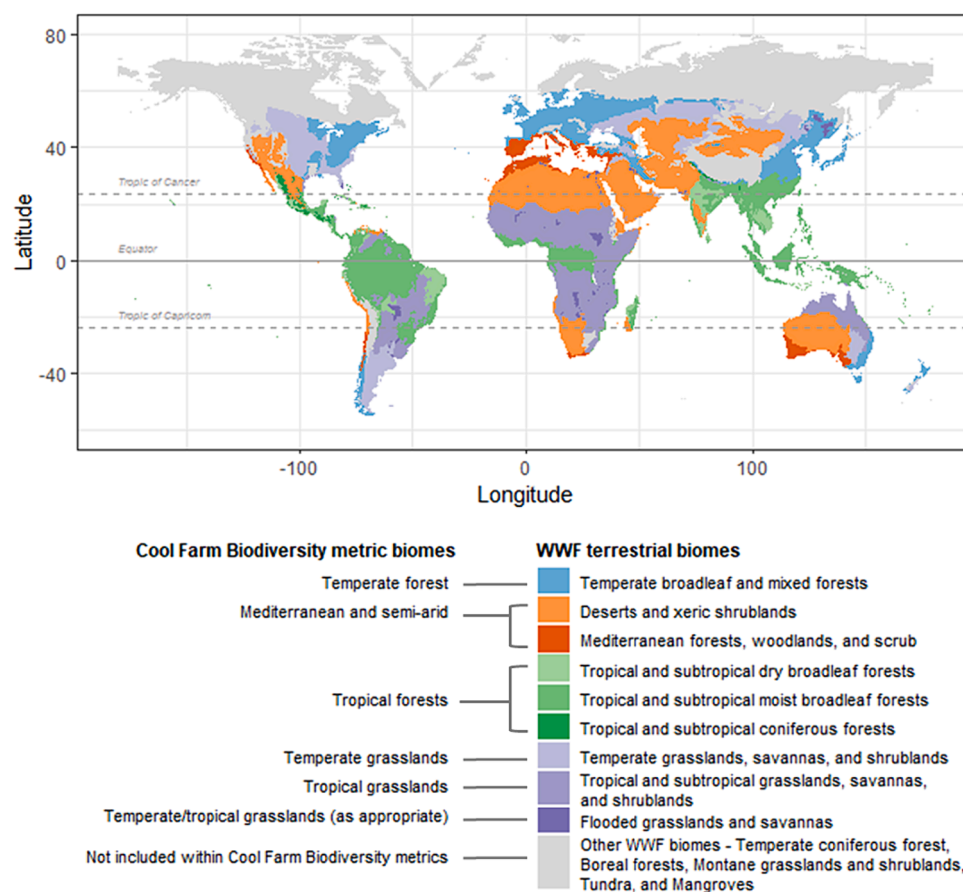
Fig. 1. Extents of nine terrestrial biomes, which together produce the majority of the world's food. Biomes that have similar types of agriculture are combined to create five Cool Farm Biodiversity metric biomes, shown in similar colours and by brackets within the legend. Biomes not included in future plans for the metric are shown collectively in grey. The map is constructed using biome data from Dinerstein et al. (2017) and downloaded from 'Resolve Ecoregions' (2020).

be effective at supporting biodiversity, either generally (section 2.2), or for a particular species group (section 2.3). Secondly 'evidence assessment' where, following the methodology of Conservation Evidence, experts review a structured summary of experimental tests of an action's effectiveness for conserving biodiversity, and score actions for certainty, effectiveness and harms, following a modified Delphi process.

Evidence assessment by experts is clearly a more rigorous approach than standalone expert judgement. The underlying database of evidence sets a high evidential standard, in that only experimental evidence is included, rather than modelling results or correlative evidence (i.e., studies that examine associations between biodiversity and habitat features without a clear link to a management action). Nonetheless, using evidence assessment exclusively would result in a narrow set of actions receiving a positive score, as many actions which may be effective for biodiversity enhancement in farmland have not been experimentally tested. Through the judgement scores, experts can positively score actions as effective, based on their technical, experiential knowledge of the effects of agricultural practices on biodiversity, or based on the balance of non-experimental evidence, such as modelling and correlative studies. In this way, the overall Cool Farm Biodiversity metric scores reflect a combination of scientific and technical, experiential or localised farming knowledge. Evidence scores have more weight than judgement scores overall, representing two-thirds of the maximum general biodiversity score available per action, as explained in section 2.2.1.

2.1.3. Biomes

Conservation strategies have differing levels of success and relevance in different places, and so tools to help support conservation efforts need

to be location specific. However, the degree of location-specificity must be balanced against the practical consideration of resources available to develop multiple versions of tools. Developing tools at the level of major habitat regions strikes a good initial balance, whilst still allowing more location-specific tools to be developed in addition, as resources allow (e. g., Brandt et al., 2018). We use 'terrestrial biomes' to define major habitat regions, as these are already spatially defined and justified in the literature.

The world is divided into 14 terrestrial biomes, each of which is an area of the world with similar environmental conditions, habitat structure, and biodiversity (Fig. 1, and 'WWF Ecoregions' 2022). Biomes are derived by grouping similar ecoregions from different parts of the world, with ecoregions defined as a "large unit of land or water containing a geographically distinct assemblage of species, natural communities, and environmental conditions", and determined following extensive literature review and collaboration with regional experts (Dinerstein et al., 2017; Olson et al., 2001; "WWF Ecoregions," 2022). The majority of the world's agricultural production occurs across nine of the fourteen biomes (temperate broadleaf and mixed forests; deserts and xeric shrublands; Mediterranean forests, woodlands, and scrub; tropical and subtropical dry broadleaf forests; tropical and subtropical moist broadleaf forests; tropical and subtropical coniferous forests; temperate grasslands, savannas, and shrublands; tropical and subtropical grasslands, savannas, and shrublands; and flooded grasslands and savannas) (Garibaldi et al., 2021). The remaining five (temperate coniferous forest; boreal forests; montane grasslands and shrublands; tundra; and mangroves) were excluded from the initial plans for tool development (Fig. 1).

Within the nine agriculturally important biomes, some are similar to
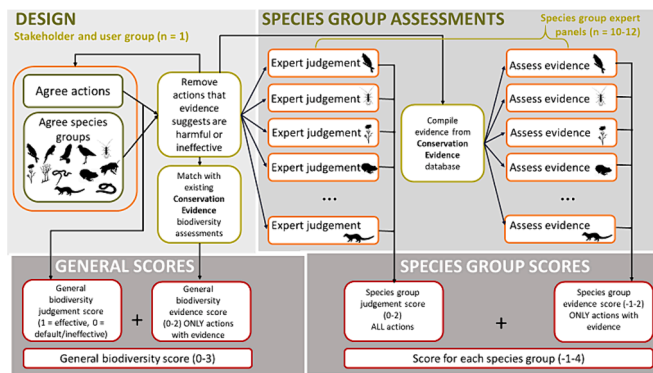
Fig. 2. Process to create the Cool Farm Biodiversity metric for a given biome. Upper panels show activities conducted by the 'Stakeholder and user group' (upper left; orange box), 'Species group expert panels' (upper right; orange boxes) and the core design team (green boxes). Lower panels show the elements of General (lower left; red boxes) and Species Group (lower right; red boxes) scores associated with each action, and how they derive from the activities (black arrows). The range of possible scores for each element is given in brackets. For both general biodiversity and species groups, the judgement and evidence scores are summed to create an overall score for the action. These scores are added together across all actions within a component (see section 2.1.4), or for a species group, to form the output scores. See text for a detailed description of the process. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

one another in the types of agriculture they support, in that the same crops are produced across more than one biome. For example, grapes and almonds are frequently grown in both 'Mediterranean forests, woodlands, and scrub' and 'Deserts and xeric shrublands' areas, usually supported by irrigation in the latter. We therefore combine these as a simplified 'Mediterranean and semi-arid' biome for the purposes of the Cool Farm Biodiversity metric. We have also combined tropical and subtropical dry broadleaf forests, moist broadleaf forests, and coniferous forests as a 'tropical forests' biome. We thus define five 'Cool Farm Biodiversity metric biomes': (1) temperate forests; (2) Mediterranean and semi-arid; (3) tropical forests; (4) temperate grasslands; and (5) tropical grasslands (Fig. 1). Two of these biomes have been completed so far (temperate forest, and Mediterranean and semi-arid). Development of another (tropical forests) is currently underway, and the remaining biomes are scheduled for later development.

2.1.4. Components

Actions that users can select are assigned to four 'components', representing different aspects of farm management. These comprise: 'Products', actions that enhance the diversity of crops and livestock (sometimes called 'agrobiodiversity') and their effects on biodiversity at a farm scale; 'Production practices', actions that relate to conservation and agronomic activities undertaken on the areas of a farm used for production; 'Small habitats', actions that involve the creation, maintenance and management of habitats in parcels of less than one hectare not used for production; and 'Large habitats', actions that involve the creation, maintenance and management of habitats in parcels of more than one hectare not used for production.

2.2. Design phase

The complete design process for the Cool Farm Biodiversity metric is summarised in Fig. 2.

At the start of development for each biome, a 'stakeholder and user group' is convened (Fig. 2). We use a purposive sampling approach to identify relevant individuals from the Cool Farm Alliance network of members, internet searches, and personal contacts. Each biome's stakeholder and user group comprises farmers, supply chain managers,

biodiversity conservation practitioners and researchers, all of whom work in the focal biome. We aim to include as a wide a range of stakeholders as possible with a diversity of backgrounds and expertise. Full lists of stakeholders, and their affiliations, for the currently available 'temperate forest' and 'Mediterranean and semi-arid' versions of the Cool Farm Biodiversity metric are given in Supplementary Tables S1-S4. For each biome separately, over the course of 1–2 days of facilitated meetings and discussion, a list of 10–12 species groups and a final shortlist of actions to be assessed are decided (Fig. 2).

These discussions consider the overall design of the tool, with a focus on the management questions and answers (actions) to be included, and the aspects of biodiversity, or 'biodiversity targets' (species groups) for which scores will be provided as outputs of the metric (Fig. 2). Actions are included on the basis that, in the expert judgement of the group, they are likely to be effective for the conservation of some component of biodiversity and are undertaken on some farms within the biome.

The temperate forest workshops took place on 2 and 9 June 2016. The group comprised six participants, of which: two were researchers acting as facilitators, three were from industry (one farmer and two supply chain managers) and one was an expert in agroecology in the biome. The group adapted an existing tool, the Gaia Biodiversity Yardstick (Kloen, 2014), which had been produced in a similar participatory process between experts and specialists from industry.

The Mediterranean and semi-arid workshop took place on 7 June 2019. The group comprised 14 participants, of which four were researchers acting as facilitators, four were from industry (two farmers and two supply chain managers), and six were experts in agroecology in the biome (four practitioners and two researchers).

Actions are removed where an evidence assessment in a published biome-relevant Conservation Evidence synopsis (the following were used for the biome versions presented here: Williams et al., 2012; Key et al., 2013; Wright et al., 2013; Shackelford et al., 2017; Dicks et al., 2013; Berthinussen,Richardson,and Altringham, 2020) has found the action either to be harmful or ineffective for the conservation of biodiversity in general (i.e., categorised as 'Likely to be ineffective or harmful' in the Conservation Evidence database; see Table 1). Discussions continue iteratively until the stakeholder and user group are satisfied with the list of actions, at which point the design phase is complete (Fig. 2).

A very important aspect of this stage is to remove 'double counting', so that each action only appears in one place in the tool, in a part of the questionnaire that is accessible to all possible users who might want to implement that action. All stakeholders are kept informed about any decision-making following the workshop, and about the progress of development of the tool, with multiple opportunities to provide feedback.

2.2.1. Assignment of 'general biodiversity scores'

Each action is assigned a score of 1 for its general biodiversity judgement score to reflect that the stakeholder and user group judged it to have a benefit for biodiversity. In some cases, the judgement score for an action accumulates scores from one or two other actions nested within it, so the user may score 2 or 3 judgement points for a single action. For example, a farm with high crop diversity – growing 'more than seven types of crop' – receives a judgement score of 3 for general biodiversity, acquiring single points for 1–3, 4–6 and > 7 types of crop. The answer options in these cases are mutually exclusive.

Actions that match with an action in a Conservation Evidence synopsis (Williams et al., 2012; Key et al., 2013; Wright et al., 2013; Shackelford et al., 2017; Dicks et al., 2013; Berthinussen,Richardson, and Altringham, 2020), and where evidence assessment has supported a 'Likely to be beneficial' or 'Beneficial' category, are assigned a 1 or 2, respectively, for their general biodiversity evidence score (Table 1). If an action has been assessed in more than one Conservation Evidence synopsis, general farmland biodiversity assessments take precedent over more focused assessments and the most recent is used. Judgement scores

and evidence scores are summed to produce the general biodiversity score for each action (Fig. 2).

2.2.2. Actions that define farm structure for comparisons and benchmarking

Some actions are included that do not affect the output of the metric and are assigned a score of 0. Actions that do not receive a score include:

- Default actions against which other actions are considered to be effective in comparison (e.g., 'none of the above' or 'conventional tillage');
- Actions included for the tool's logic, to specify which other actions are available to the focal farm (e.g., whether a farm has annual field crops, so that exclusively pasture or perennial crop farmers do not have to answer questions that do not apply to their farms); and
- Actions that specify or provide information about the farming or landscape context, so that scores for similar types of farms can be compared in any subsequent benchmarking process. These include options appropriate to the biome, such as types of agricultural product or farm business, irrigation, and landscape type. This enables users working across multiple farms, at the level of an entire supply chain for example, to compare scores among similar types of farm, for whom the same set of options are likely to be available.

2.3. Species group assessment phase

2.3.1. Assignment of 'species group scores'

For each biome, a panel of experts for each species group is convened. Experts are academic and NGO-based researchers who have worked on the focal species group, within that specific biome, found using their publication histories. Experts provide input through two steps: A first 'expert judgement' step, and secondly an 'evidence assessment' step (Fig. 2).

For temperate forest, twenty-seven researchers were recruited, all with a publication history of agroecological research in temperate regions (Supplementary Table S2). For Mediterranean and semi-arid, twenty-three researchers were recruited, all with a publication history of agroecological research in one or more of the regions covered by the biome (Supplementary Table S4). Every continent with Mediterranean and semi-arid regions except Australia was represented in the publication histories of the experts. We included experts working in the Caatinga biome of north-east Brazil, which was previously classified as 'Deserts and Xeric Shrublands' (Olson et al., 2001), but whose classification was changed to 'Tropical and Subtropical Dry Broadleaf Forests' in the most recent iteration of the biome boundaries (Dinerstein et al., 2017), indicating some uncertainty about classification specific to this ecoregion.

To complete the first 'expert judgement' step, each expert panellist is provided with a list of the actions agreed in the design phase, with definitions where appropriate. Experts score each action, using their expert judgement, as either: ineffective (0), effective (1), or critical to the conservation of their species group on farmland in the focal biome (2). Within each species group panel, the median score across experts forms the expert judgement score, rounded down to the nearest integer. To complete the 'evidence assessment' step, the Conservation Evidence database of studies is searched for experimental tests of the effectiveness of each action at conserving each species group on farmland within the focal biome. These are collated and summarised, following Sutherland et al (2019). Each summary is assessed by the expert panel, via an online portal following the procedure used by Conservation Evidence (Sutherland et al., 2019). Actions are assigned evidence scores according to the final effectiveness categories, as shown in Table 1: likely to be ineffective or harmful (-1), unlikely to be beneficial/unknown/trade-off between benefits and harms (0), likely to be beneficial (1), beneficial (2).

Expert judgement scores and evidence scores are summed to produce

the species group score for that action (Fig. 2). Thus, for species group scores, evidence and judgement are given equal weight. This is appropriate because at the level of species groups, it is possible to provide information about the magnitude, or importance, of the effects of each action. Scoring the magnitude of effects is very challenging for general farmland biodiversity, because different groups often respond differently. Our method of evidence assessment does not currently allow for effect sizes, although this is likely to become possible in the future, as meta-analytic approaches to large-scale evidence synthesis become widely established (Shackelford et al., 2021). To compensate for a lack of effect sizes in the evidence, we assign more weight to expert judgement of the effects on species groups.

2.4. Actions that mitigate harm

Some actions are included in the metric not because they benefit biodiversity, but because they mitigate harm to biodiversity, arising from some other farming practice, e.g. actions that aim to minimise the biodiversity impacts of using crop protection products. The metric handles these actions by changing how a user's answers contribute to their farm's score. Specifically, a harmful action, for example using insecticides, results in the farm losing score compared to the default option of not using insecticides. Then, the set of actions which mitigate that specific harm result in the previously lost score being regained, with each action contributing part of the lost score. This is calculated by:

$$\widehat{m_i} = h \times \frac{m_i}{\sum_{i=1}^{|M|} m} \tag{1}$$

where h is the score lost due to the harmful action, and m_i is the returned score for one of a set of i specific mitigating actions m. The total score available for taking the full set of mitigating actions $|M|$ is therefore equal to the harm score (h), and the score awarded for each mitigating action ($\widehat{m_i}$) is said to be 'normalized' by the harm score. Thus, mitigation cancels out harm. Both the harmful action and the mitigating actions are scored according to expert judgement like any other, but expert panellists are asked to score the actions for whether they are harmful or effective at mitigating, as appropriate, on a scale of 0–2. In most cases, the harmful actions are crop protection chemicals used to target specific groups of pests (such as insects, fungi and other diseases, weeds), and the harm score is set at 2, so that individual mitigating actions have some noticeable impact on the score.

Following the example of mitigating insecticide use, assume that experts judged insecticides to be harmful ($h = 2$), and identified four mitigating actions (m_1, m_2, m_3, m_4), each judged to be equally effective (i.e. each has a score of 1). A user that selected two of those mitigating actions would therefore regain one point overall, with the following calculation, from Equation (1) returning a score of 0.5 for each mitigating action:

$$2 \times \frac{1+1}{1+1+1+1} = 1$$

This allows for the metric to be responsive to farmers using a defined set of Integrated Pest Management (IPM) techniques to minimise harms, but only rewards actions that expert panellists have judged to be effective at mitigating harms. This design means farmers taking all the recommended mitigating actions will receive the same score as farmers not using chemical crop protection products at all. This position might be challenged, since organic farms are well known to host more species (i. e., more biodiversity) at a field scale (Tuck et al., 2014), although the effect is not replicated across all taxa, and is smaller at the whole farm-scale at which the Cool Farm Biodiversity metric operates (Schneider et al., 2014). However, the metric is designed to drive improvement, rather than to make predictions about actual biodiversity outcomes. Our design allows the largest number of farmers to demonstrate improvements, irrespective of their organic status, in a broad range of

agricultural systems.

A consequence of this design is that when you first enter the tool, there are existing scores on the scoreboard, because you have not yet selected the harmful actions, which subtract these scores. In the Temperate forest version of the metric, this logic for mitigating harms was only originally used for general biodiversity scores, and not for species group scores. In the Mediterranean and semi-arid version, added after several years of testing, and on-line user feedback, the species groups also have scores for harms and mitigating actions. This difference remains because we prioritise expansion of global coverage through development of new biomes over design improvements to existing biomes, guided by the user community.

2.5. Calculating scores per component or species group

The metric calculates a score for each of the four components (Products, Production Practices, Small Habitats and Large Habitats), using the general biodiversity scores, and for each species group across all components, using the species group scores. A user's overall score for a component ($Score_{comp}$) is the sum of achieved general biodiversity scores ($Score_Q$) across all questions in the component, as a percentage of the maximum possible general biodiversity score across all questions in the component (Equation (2)). The achieved score for an individual question, ($Score_Q$) is the sum of the general biodiversity scores of all actions selected by the user ($a_{selected}$, Equation (3)). The maximum general biodiversity score available for a question ($Score_{QMax}$) is the sum of all positive action scores for which multiple answers are permitted ($a_{multipleanswers}$), and the highest available score from any subset of answers from which only one can be chosen ($a_{mutuallyexclusive}$) (Equation (4)). A user's score for each species group ($Score_{SG}$) is the sum of achieved scores for that species group, based on actions selected in each question ($Score_{sgQ}$), expressed as a percentage of the maximum possible score for that species group across all components ($Score_{sgQMax}$) (Equation (5)). $Score_{sgQ}$ and $Score_{sgQMax}$ are calculated following Equations (3) and (4), but using scores for the relevant species group, instead of general biodiversity scores.

$$Score_{comp} = \frac{\sum Score_Q}{\sum Score_{QMax}} \times 100 \tag{2}$$

$$Score_Q = \sum a_{selected} \tag{3}$$

$$Score_{QMax} = \sum a_{multipleanswers} + max\{a_{mutuallyexclusive}\} \tag{4}$$

$$Score_{SG} = \frac{\sum Score_{sgQ}}{\sum Score_{sgQMax}} \times 100 \tag{5}$$

2.6. Treatment of scale

Users input the areas of different types of habitats on their farms,
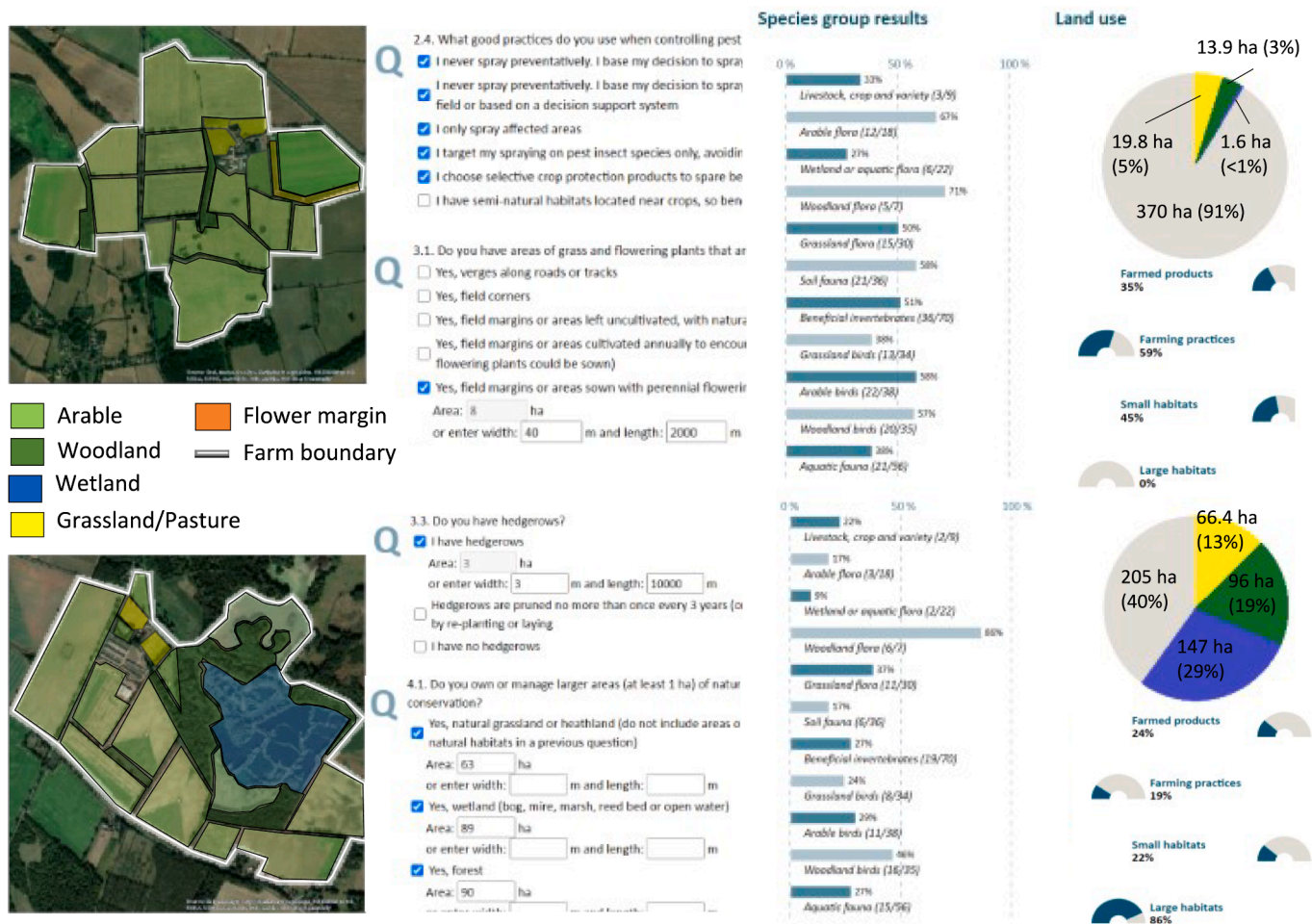
Fig. 3. Example of user data input and outputs for two hypothetical farms in the temperate forest biome. Left - hypothetical farms, one with minimal natural land cover but a large number of actions effective for biodiversity conservation (top), another with significant areas of natural land cover, but a minimal number of actions that are effective for biodiversity conservation (bottom); Centre - short excerpts of the user data input for the two hypothetical farms (illustration only: see Supplementary Figure S1 for a more detailed view of the data input screen); Right - results returned by the metric for the two hypothetical farms.

within the small habitats and large habitats components. These broad habitat types are defined, in biome-appropriate terms, in the questions with which users are presented. Users have the option of entering areas directly, in hectares, or entering the dimensions, in metres. The areas of these habitat types are reported in the metric's results output, including as a percentage of total farm area (Fig. 3). Total farm area, including cropped and uncropped areas, is a separate input.

Currently, these areas do not affect the score a farm receives for the respective components and are reported independently in the metric's output. This is because, although larger patches of habitats are well known to support more species across multiple taxa (Connor and McCoy, 1979), it is very difficult to set thresholds or scales for these across the geographical scope of the metric. However, by collecting and reporting these areas, the metric retains the possibility of implementing an adjustment or scaling of the general biodiversity scores for small and large habitats, should evidence relating to a sufficient geographic extent become available (*cf.* Meixler, Fisher, and Sanderson, 2019).

2.7. Presentation and interpretation of results

Users are presented with their results in a dynamic view that updates as they input their actions. These comprise: 1) General biodiversity scores relative to the maximum possible (expressed as a percentage and structured by components); 2) Species group scores relative to the maximum possible (expressed as a percentage); 3) A breakdown of their on-farm land cover by broad habitat types (expressed in hectares and as a percentage). A screenshot of the Cool Farm Biodiversity metric data entry page is provided in Supplementary Fig. S1. A hypothetical set of results to illustrate two different farms in the temperate forests biome is shown in Fig. 3.

As detailed above, the general biodiversity and species group scores represent what proportion of effective actions are undertaken as part of a given farm's management, with extra weight given to actions well-supported by evidence. They are not expected to predict actual biodiversity outcomes (i.e. species population densities, richness or diversity on the ground), either for the species groups or biodiversity in general. This is because biodiversity responses to management are highly context-dependent, and scale-dependent, including a dependence on processes that operate at scales larger than the farm. Predicting actual biodiversity outcomes of farm management, at farm scale, using a scoring system that operates at biome scale, is very unlikely to be reliable or accurate.

Moreover, the scores are only intended to be comparable between farms of a similar type (see section 2.2.2). By way of example, consider two farms, one arable only and another with a mix of arable, pasture and perennials; the latter farm has the opportunity to score higher as they will likely undertake a wider range of actions across their different operations.

2.8. Example of use

To demonstrate that the metric is responsive to relatively minor changes in farming practice, we present outputs from a typical farm in the temperate forest biome before and after changing farming practices to benefit biodiversity. We use a large arable farm in the UK that was initially under an agri-environment scheme (AES) called 'Entry Level Stewardship' before also joining the higher level 'Mid-Tier Stewardship' scheme. Entry Level Stewardship was open to all farmers in England and Wales and typically required farmers to make only small, if any, changes to their practice (Hodge and Reader, 2010), whereas Mid-Tier Stewardship is a competitive AES in which farmers are rewarded for more costly actions that benefit biodiversity. Funding is limited based on farm area, and farmers are only funded for actions that correspond to regional priorities (Franks, 2019). This means that we would expect the demonstration farm to have undergone relatively minor improvements in farming practice for biodiversity, and therefore if the metric is responsive and sufficiently sensitive, the outputs should show a small increase in the general biodiversity and species scores. This example is from a commercial farm using the Cool Farm Biodiversity metric. We were granted access to the output scores, but not the details of the farm location, agri-environment management agreements, or inputs to the metric.

3. Results

Here, we present results from the 'temperate forest' and 'Mediterranean and semi-arid' versions of the metric.

3.1. Design and species group assessment

The full set of scores underlying the temperate forest and Mediterranean and semi-arid versions of the Cool Farm Biodiversity metric are provided in the Supplementary Information Table S12.

3.1.1. Temperate forest biome

The stakeholder and user group defined 150 actions, of which 115 received a positive general biodiversity judgement score. One action, 'Reduce grazing intensity on grassland', was removed because it was assessed as 'Likely to be ineffective or harmful' in the Conservation Evidence database. The actions were structured as answers to 29 questions for users of the tool to answer. The species groups were defined mostly on the basis of habitat associations (see Table 2). Eleven species groups were defined, see Table 2 and Supplementary Tables S5-S9 for their definitions.

Overall, 23 (20 %) out of 115 actions received an evidence score of 1 or more, either for general biodiversity (Table 2A), or for one or more species groups. Among these actions supported by evidence, four received a positive score for one or more species groups and not for general biodiversity.

The actions supported by evidence (therefore those with the highest scores) were mostly placed in the 'production practices' and 'small habitats' components and involved either creating in-field habitats (e.g., overwinter stubbles, skylark plots), reducing inputs by switching to sustainable alternatives (e.g., reducing pesticides, adding organic matter, reducing or eliminating soil tillage) or managing field margins.

Across the temperate forest species groups, different numbers of actions received positive judgement scores, indicating that they were thought to be effective, or critical, for the conservation of that species group: Livestock crop and variety, 16; Arable flora, 11; Wetland and aquatic flora, 14; Woodland flora, 5; Grassland flora, 19; Soil fauna, 26; Beneficial invertebrates, 47; Grassland birds, 27; Arable birds, 28; Woodland birds, 26; Aquatic fauna, 40. For full details of which actions contribute to the conservation of each species group in the temperate forest biome see Supplementary Table S10. These sets of actions can be used as a starting point for Biodiversity Action Plans focused on particular target species groups.

3.1.2. Mediterranean and semi-arid biome

The stakeholder and user group workshop defined 188 actions of which 148 received a positive general biodiversity judgement score. The actions were structured as answers to 30 questions for users of the tool to answer. The workshop participants took the view that defining species groups by habitat use (e.g., as in the temperate forest biome) was difficult to apply across the biogeographic range of the biome and instead used functional traits (e.g., feeding guilds) to delineate which taxa species groups referred to. Twelve species groups were defined, see Table 2 and Supplementary Text S1 for their definitions.

Overall, 19 (13 %) out of 148 actions received an evidence score of 1 or more either for general biodiversity, or for one or more species groups (Table 2B). Relative to the temperate forest biome, fewer actions for the Mediterranean and semi-arid biome received evidence scores for general biodiversity. Of the 19 actions that received any evidence score, all

Table 2

Scores achieved by the highest scoring, evidence-supported actions across both currently available biomes. In the temperate forest biome (A), actions are included here if they received a positive general biodiversity evidence score. In the Mediterranean and semi-arid biome (B) actions are included here if they received a positive evidence score for general biodiversity or two or more species groups. Column headings are defined as follows: 'Action', short description of action; 'General biodiversity', evidence scores and judgement scores assigned to action in the design stage (see Fig. 2); 'Species groups', total scores assigned by species group expert panels (evidence + judgement). Number in parentheses shows the contribution to the total score made by the 'Evidence scores', when available. Links to the relevant evidence are given in Table S12. Absence of a number in parentheses for species group scores indicates that the score was based on expert judgement alone.

A) Temperate forest biome

| Action | General biodiversity | | Species groups | | | | | | | | | | |
	Evidence	Judgement	Crop/livestock variety	Arable flora	Wetland or aquatic flora	Woodland flora	Grassland flora	Soil fauna	Beneficial invertebrates	Grassland birds	Arable birds	Woodland birds	Aquatic fauna
Use no pesticides	2	2	0	2(2)	0	0	1(1)	0	2(2)	0(0)	1(1)	0(0)	0
Aim to reduce pesticide use	2	1	0	2(2)	0	0	1(1)	0	2(2)	0(0)	1(1)	0(0)	0
Reduced tillage	1	1	0	1(1)	0	0	0	3(2)	1(1)	0(0)	0(0)	0	0
No tillage	1	1	0	2(1)	0	0	0	4(2)	2(1)	0(0)	1(0)	0	0
Add organic matter (manure)	2	1	0	0(0)	0	0	0(0)	4(2)	1(1)	1	1	1	0
Add organic matter (compost)	2	1	0	0(0)	0	0	0(0)	4(2)	1(1)	1	0	0	0
Overwinter stubble	1	1	0	1(0)	0	0	0	1	0(0)	0	3(1)	1	0
'Areas of cereal fields cultivated, unsown ('Skylark plots')	2	1	0	4(2)	0	0	1	0	2(1)	2(2)	2(0)	2	1
No slurry or mineral fertiliser in grass fields	1	1	0	0	1	0	3(1)	2(1)	1(0)	2(1)	0(0)	0(0)	1
No spring mowing/grazing in grass fields	1	1	0	0	0	0	2(1)	0	0(0)	2(1)	0	0	0
Unmown strips in grass fields	1	1	0	0	0	0	0	0	1(0)	2(0)	0(0)	0(0)	1
Field margins (uncultivated)	2	1	0	0(0)	0	0	1(-1)	0	3(2)	1(0)	1(0)	1(0)	1
Field margins (annually cultivated)	2	1	0	4(2)	0	0	1	0	2(1)	2(2)	2(0)	2	1
Field margins (sown flowers)	2	1	0	0(0)	0	0	2(1)	0	3(2)	0(0)	2(0)	2	1
Field margins (sown perennial grasses)	2	1	0	−1(-1)	0	0	2(1)	0(0)	3(2)	0(0)	2(1)	1(0)	0(0)
Field margins (sown bird seed)	2	1	0	1	0	0	1	0	2(1)	0	4(2)	4(2)	1
Hedgerows managed for wildlife	1	1	0	0	0	2(1)	0	0	1(0)	0(0)	0(0)	1(0)	1
Watercourses (manage native bank vegetation for wildlife)	1	1	0	0	2	0	0(0)	0	1(0)	1(0)	0(0)	0(0)	4(2)
Provide nest boxes of owls or birds of prey*	2	1	0	0	0	0	0	0	0	2(2)	2(2)	1	0

B) Mediterranean and semi-arid biome

| Action | General biodiversity | | Species groups | | | | | | | | | | |
	Evidence	Judgement	Pollinators	Predatory Invertebrates	Insectivorous birds and bats	Fruit and seed eating birds	Wading birds	Birds of prey	Reptiles	Carnivores	Soil fauna	Aquatic fauna	Native scrub plants	Native wetland plants
No tillage (arable)	–	1	2(2)	1(0)	1	1	0	1	1	0	3(1)	1	2	1
Cover crops (perennial fields)	0	1	1	2	2(1)	2(1)	0	1	1	1	4(2)	1	1	1
Native ground cover (perennial fields)	0	1	1	2	2(1)	3(1)	0	2	2	1	3(2)	1	2	1

Table 2 (*continued*)

B) Mediterranean and semi-arid biome

Action	General biodiversity Evidence	General biodiversity Judgement	Species groups Pollinators	Predatory Invertebrates	Insectivorous birds and bats	Fruit and seed eating birds	Wading birds	Birds of prey	Reptiles	Carnivores	Soil fauna	Aquatic fauna	Native scrub plants	Native wetland plants
Add organic matter (compost)	1	1	0	0(0)	0	1	1	1	0	1	4(2)	0	2(1)	1
Plant flowers (perennial fields)	0	1	4(2)	4(2)	1	2	0	1	2	1	1	1	2(0)	0
Field margins (annual cultivation)	0	1	3(2)	4(2)	2	2	1	1	2	2	1	1	2	0
Field margins (sown flowers)	0	1	4(2)	4(2)	2	2	1	1	2	2	2	1	2	0
Hedgerows (presence)	0	1	3(2)	3(1)	2	2	0	2	2	2	1	1	2(0)	0
Watercourses (restore native vegetation)	1	1	2	2	3(1)	3(1)	2	2	2(0)	2	1(0)	2(0)	2(0)	2(0)
Reduce grazing frequency (large natural habitats)	0	1	0	1	2	3(1)	1	2	1	1	1	1(0)	3(1)	2
Exclude grazers (large natural habitats)	0	1	0	1(0)	3(1)	3(1)	3(2)	3(1)	1(0)	0(0)	2(0)	2(1)	3(1)	2

*Barn owls were classed as 'grassland birds', kestrels classed as 'arable birds', during the species group evidence assessment (see Supplementary Information Table S6).

received a score for one or more species groups, of which just two actions received an evidence score for general biodiversity. Actions that received evidence scores were mostly placed in the 'production practices' or 'small habitats' components, and were similar to those supported by evidence in the temperate forest biome. In addition to these actions, cover crops and ground cover in different productive settings also received evidence scores for multiple species groups. The Mediterranean and semi-arid biome had a greater focus on restoring natural habitats and vegetation with evidence scores for several actions related to ceasing and reversing the impacts of overgrazing as well restoration of native vegetation around watercourses.

Across the Mediterranean and semi-arid species groups, different numbers of actions received positive judgement scores, indicating that they were thought to be effective, or critical, for the conservation of that species group: Pollinators, 62; Predatory invertebrates, 83; Insectivorous birds and bats, 114; Fruit and seed-eating birds, 129; Wading birds, 98; Birds of prey, 132; Reptiles, 89; Carnivores, 92; Soil fauna, 98; Aquatic fauna, 68; Scrubland plants, 81; Wetland plants, 99. For full details of which actions contribute to the conservation of each species group in the Mediterranean and semi-arid biome, see Supplementary Table S11.

3.2. Examples of use

As of 27 April 2023, a total of 4,355 individual farm assessments had been made using the Cool Farm Biodiversity metric, representing farms in 105 different countries (pers.comm, Cool Farm Alliance). We cannot see details of these assessments, because data inputs to the Cool Farm Biodiversity metric belong to the users, most of whom are commercial farms, suppliers or consultants. We were granted access to one set of output data from a UK farm, to illustrate the sensitivity of the metric.
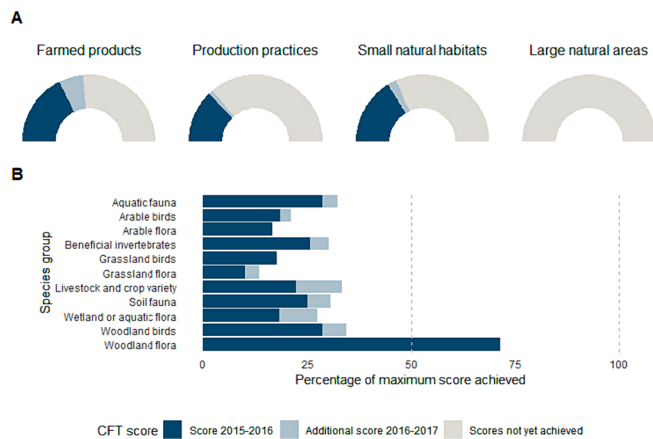
Following adoption of an agri-environment scheme, small improvements were made in the general biodiversity scores across three of the four components (Products, +11.8 %; Production practices, +2.1 %; Small habitats, +4.1 %; See Fig. 4). The 'large habitats' component score did not improve, since the farm had no qualifying large habitats (>1 ha in a single parcel) and did not create any. Across the species groups relevant to the biome (temperate forest), the example farm improved its score for eight out of eleven species groups (Aquatic fauna, +3.6 %; Arable birds, +2.6 %; Beneficial invertebrates, +4.3 %; Grassland flora, +3.3 %; Livestock and crop variety, +11.1 %; Soil fauna, +5.6 %; Wetland or aquatic flora, +9.1 %; Woodland birds, +5.7 %; See Fig. 4).

4. Discussion

Here we present a metric that allows farmers or agri-food companies working with networks of supplier farms, to monitor, benchmark and report improvements in conserving and managing biodiversity at farm scale. The Cool Farm Biodiversity metric is unique, because unlike alternative available software tools (e.g., Gaia Biodiversity Yardstick (Kloen, 2014), the European Biodiversity Performance Tool (EU LIFE Initiative, 2021) and NatCapMap ("NATCAP MAP," 2022), its scores and outputs are designed to be based on evidence. Its action-based, farm-scale approach is designed to increase usability, and it can be adapted to the different ecological and agronomic conditions found across biomes, meaning it is adaptable to all of the world's major food producing regions. By way of a real-world example, we demonstrate that the metric is responsive to even minor changes in farming practice, and thus gives a useful indication of how 'biodiversity-friendly' the suite of current management practices on a farm are expected to be.

Several other initiatives have used a similar points-based approach to develop or evaluate national or European accreditation schemes for biodiversity on farms (Birrer et al., 2014; Gabel et al., 2018; Schader et al., 2014; Tzilivakis et al., 2016; van Doorn, Anne; Jongeneel, Roel, 2020; Zellweger-Fischer et al., 2016), or as part of wider sustainability assessments, as in the SMART-Farm tool (Schader et al., 2016). To our

Fig. 4. Outputs for an example farm before (2015–2016) and after (2016–2017) adoption of an agri-environment scheme that required a higher level of management intensity for biodiversity, showing general biodiversity scores (A) and species group scores (B). CFT = 'Cool Farm Tool'.

knowledge, none of these are freely available in the form of easy-to-use online software.

The combination of stakeholder priorities, expert judgement and evidence synthesis that underpins the scoring system in the Cool Farm Biodiversity metric allows it to incorporate technical agronomic knowledge, biome-specific priorities (i.e. different actions and species groups are prioritised for each biome), and importantly, gives stakeholders across the agri-food sector a voice in the governance of biodiversity, which can be expected to lead to better outcomes for biodiversity in the long term (MacLeod et al., 2022).

We acknowledge that, as with any self-assessment tool designed for large-scale industry use, there is a risk that inputs do not reflect management on the ground, or biodiversity outcomes. Our purpose is partly to support decision-making at farm scale and to drive improvement in practice. It is the responsibility of organisations using this metric for biodiversity reporting or certification to incorporate auditing processes. Some aspects of the data, such as the habitat areas, can potentially be checked using remote sensing, but others will require detailed on-farm auditing or independent assessment.

4.1. Combining 'bottom-up' and 'top-down' approaches to biodiversity management

The Cool Farm Biodiversity metric focuses on relatively small-scale, localised actions, that most farmers could take on their land. These include aspects of farmed products (e.g., diversity of crops and livestock), production practices (e.g., use of cover crops, organic fertilisers, or agroecological pest control), small habitats (e.g., wildflower strips, conservation areas on steep slopes), and large habitats (e.g., large areas of the farm set aside for nature). There is clear evidence – much of which has been used to develop the metric – for the benefits of a range of these on-farm management actions for local biodiversity (for example: Dicks et al., 2013; Birrer et al., 2014). Furthermore, this small-scale approach means the actions are easily achievable for many farms and landscapes, giving potential for widespread uptake.

We acknowledge that a tool designed to be easy to use for farmers and applicable to all possible farm structures and types across an entire biome, is still relatively coarse at this local scale. The actions are those expected to benefit biodiversity across a range of farming contexts. The tool is unlikely to include all actions that might benefit biodiversity in a particular context. For example, targeted actions for local endemic species with restricted ranges may not be captured, nor would actions very specific to farming systems that are not widely distributed. Scores should not be compared across very different farming systems. We

recommend that users only compare farms of similar type, using non-scoring actions that specify or provide information about the farming or landscape context, as explained in section 2.2.2.

In addition to this localised 'bottom-up' approach, larger 'top down' actions that seek to limit and avoid habitat loss at landscape or regional scales, are crucial to protect biodiversity (Watson et al., 2014). A large number of species are unable to persist in agricultural landscapes and depend on the maintenance of large areas of natural habitat (Gibson et al., 2011). Examples of rigorous 'top-down' approaches already in use by industry include 'Science based targets' ("Science Based Targets Network," 2022) and the 'No net loss agenda' (Bull et al., 2013; Simmonds et al., 2020). Science-Based Targets are a means by which businesses, typically those with multiple sites, can align their biodiversity action strategy with globally agreed goals, such as those set by the Convention on Biological Diversity. For example, the post-2020 Global Biodiversity Framework (Convention on Biological Diversity, 2022) contains specific targets for the proportion of land under protected areas, and the proportion under effective restoration. These proportions can be calculated at a range of scales. The 'No net loss agenda' is an agreed standard by which businesses can mitigate and offset impacts of development projects so that they achieve no overall negative impact on biodiversity. Both Science-based Targets and the 'No net loss agenda' are generally applied at landscape scales or larger, potentially involving tens or hundreds of individual farms.

There is substantial potential to combine and integrate these bottom-up and top-down approaches to measuring biodiversity impacts. For example, an international company making chocolate products might use the tropical forests version of the Cool Farm Biodiversity metric (currently in development) to derive scores for its individual cocoa supplier farms, which could include thousands of smallholders, and use the scores to reward those following the best practice management for biodiversity, through a pricing structure. The same company might also use the Science-based Targets framework to set targets for areas of natural habitat protected, or agricultural land 'restored' (Pashkevich et al., 2022), at a larger scale, in the regions it operates.

4.2. Evidence gaps revealed

Across the two biomes for which the Cool Farm Biodiversity metric has so far been completed (temperate forest, and Mediterranean and semi-arid) only a minority of actions were supported by evidence and therefore received an evidence score (whether for general biodiversity or species groups). The evidence gaps are distributed across the components, and partly reflect a lack of focus on agricultural management interventions in the Conservation Evidence database. For example, evidence is limited on the mitigation of harms from crop protection products, various actions to protect soil health, increase diversity within crop fields or provide wildlife nesting or refuge resources within farmland (see Table S12). They also partly reflect a lack of manipulative experiments in the agroecological literature, as noted by other authors. Similar evidence gaps were found during development of a similar tool for New Zealand agriculture, where evidence assessment for two farm management actions, based on the Conservation Evidence database, found relevant evidence for only four of the 10 target biodiversity groups prioritised by stakeholders, and no relevant evidence from New Zealand itself (MacLeod,Brandt,and Dicks, 2022).

Of the two biomes, more actions were supported by evidence in the temperate forest biome than in the Mediterranean and semi-arid biome. The evidence gaps indicate a need for more experimental research into the effects of agricultural practices on biodiversity, with the difference between the biomes reflecting a geographical bias in agroecological research towards Western Europe (Dicks,Walsh,and Sutherland, 2014). As well as a likely real difference in research levels between biomes, there is also likely to be a difference in the extent to which experiments originally published in languages other than English, are synthesised into the evidence base used here. This is a bias that affects most 'global'

evidence syntheses (Amano et al., 2016; Lynch et al., 2021) and underlines a need for greater investment in synthesizing evidence from literature written in languages other than English (Nuñez and Amano, 2021).

These research and evidence-synthesis gaps mean that the Cool Farm Biodiversity metric is currently far from fully parametrised by evidence, creating a heavy weighting towards those actions for which there is strong evidence of effectiveness. However, the metric strikes a compromise between the need for evidence and the pressing need to provide conservation guidance. The design ensures that more points are awarded for the actions that we are most certain about and fewer points for actions that are less strongly evidenced, as well as a flexible structure that can be updated easily when more evidence becomes available.

4.3. How do Cool Farm Biodiversity scores relate to biodiversity in situ?

This is a common question from users of the Cool Farm Biodiversity metric. Our answer begins with a justification for the design we have chosen. The Cool Farm Biodiversity metric is designed to be accessible to all farmers, and to reward their efforts to take actions to support biodiversity. It provides equal credit to growers taking the same actions, regardless of the landscape context their farm is in, and therefore their potential to support high levels of biodiversity. By not penalising intensive farmers starting from a low biodiversity baseline, our approach encourages widespread engagement and focuses on improvement from a baseline. We also do not require any measurement of biodiversity on farms (species abundances, species or habitat diversity for example), which would advantage farmers with access to the specialist resources or expertise required to do this.

There are some downsides to this approach, including that: (1) there is no scope for scoring actions in a context-dependent way; (2) some conservation priorities, in particular the conservation of endangered or protected species, are given a relatively low emphasis in the metric because relevant actions are specific to a particular species or context, and not widely applicable; and (3) as we are not measuring farm-level outcomes for biodiversity, when comparing across farms, higher scores might not *always* be associated with higher biodiversity on the ground. The latter is particularly true because the scores awarded for Small and Large Habitats are not scaled by area, so two farms with small and large areas of the same habitat types would score the same, but likely have very different levels of biodiversity on the ground. As mentioned in section 2.6, there is scope to reflect this in the scores, based on information collected by the metric. For example, additional points could be added to the Large Habitats score if a pre-defined proportion of total farm area (e.g. 20 %) is native habitats, as recommended by Garibaldi et al (2021). Alternatively, scores for specific habitat types or species groups could be magnified if a certain level of connectivity or proportional area of a relevant habitat type is reached. This scale-sensitive scoring is an upgrade we plan for the future, when we hope there will be clearer evidence about the relationships between different biodiversity measures and habitat areas.

Regarding context-dependent scoring (points 1 and 2), consider sown flowering field margins, one of our highest scoring actions (see Tables 2A and 2B). The effectiveness of these for supporting biodiversity is known to depend on 'ecological contrast' in floral resources and the structure of the surrounding landscape, with flower strips more effective in moderately simplified landscapes than in complex landscapes (Scheper et al., 2013). Whilst it would be possible to adjust the scores according to surrounding landscape structure, this would penalise some farmers for aspects of their landscape setting that are outside their control. On the other hand, incorporating landscape structure into the scoring might incentivise farmers to work with others around them, to improve biodiversity at landscape scale. This is a development we are considering for the future. It would allow the metric to account for landscape-scale processes, such as the positive effects of higher edge density, small field size, and distributed small habitat patches on

biodiversity in farmland (Clough et al., 2020; Martin et al., 2019; Riva and Fahrig, 2023).

Consider also those species that are threatened in some areas of a biome, but invasive in others, such as *Hydropotes inermis* (Chinese water deer), invasive in parts of the United Kingdom, but classed as 'Vulnerable' by the International Union for the Conservation of Nature's Red List ("IUCN Red List," 2022) in their native range in China and Korea (Putman et al., 2021). Actions that might specifically support such species probably should not be rewarded in the invasive range, but such logic would require a level of granularity that is way beyond the design of the Cool Farm Biodiversity metric.

Despite these shortcomings, there ought to be a positive correlation between the Cool Farm Biodiversity metric scores and actual biodiversity levels measured on farm, at least in the way these two metrics change over time, if not across farms. If there is no such correlation over time, then the actions incentivised by the tool are not providing the expected improvements and could instead be an unnecessary cost to both farm businesses and overall food production. Birrer et al. (2014) validated their Credit Point System against real biodiversity across 133 Swiss farms, by showing that the points score could be used as a predictor for most measures of species richness or density. Validating the Cool Farm Biodiversity metric in the same way requires considerable investment in 'ground-truthing' studies that measure biodiversity in large numbers of farms across a range of biomes; this is a central aim of our ongoing research programme.

4.4. Conclusions and future directions

The Cool Farm Biodiversity metric is the first evidence-based on-line tool for easily assessing the biodiversity-friendly management strategies used on farms. It can be used by farmers and supply chain members across the world to give rapid, biome-specific, management recommendations and to quantify the current state of agricultural restoration for biodiversity. It allows farmers to score points to demonstrate the good they are doing, helping to incentivise engagement. Its design is stakeholder-led, supported by rigorous scientific evidence, and offers flexibility for future updates as more evidence becomes available.

Future developments are planned both to enhance the user experience and to extend the ecological information used by the tool to evaluate biodiversity management. These will include: (1) Ongoing updates as new evidence becomes available via published literature or additional ground-truthing experiments; (2) Development of a GIS mapping function into which users can enter polygons showing the extent and spatial arrangement of small and large habitat features on their farms. This will enable automatic calculation of areas and allow for scoring to reflect the value of on-farm habitats in providing connectivity between other features in the surrounding landscape – a point of key importance for maintaining biodiversity in the long-term (Hanski, 1998); (3) Development of functions to integrate the Cool Farm Biodiversity metric with broader landscape-scale industry-led sustainability approaches, such as SBTs, and the 'No net loss agenda'; and (4) Continued expansion of the metric across the remaining biomes. The Cool Farm Biodiversity metric provides a valuable sustainability tool that has already been used by over 4,000 farms worldwide, and has the potential to drive an increase in the application of biodiversity-friendly practices in agricultural areas in many parts of the world.

Software and data availability statement

Two versions of the Cool Farm Biodiversity metric – 'Temperate forest' and 'Mediterranean and semi-arid' – have been freely available online for registered users (registration is free) since 2016 and 2021 respectively. They can be accessed at https://app.coolfarmtool.org/ using only browser software. The software is managed by Cool Farm Alliance, a community interest company. Registered address: 87b Westgate, Grantham, Lincolnshire, NG31 6LE England. Email:

info@coolfarmtool.org. It was built, under contract, by Anthesis group. Technical documentation of the methods, as described in this paper, has been available to users on request since 2020. All underlying scores are provided here as Supplementary Information (Table S12).

CRediT authorship contribution statement

Liam P. Crowther: Writing – original draft, Visualization, Validation, Investigation, Formal analysis, Data curation. **Sarah H. Luke:** Writing – review & editing, Visualization, Formal analysis, Data curation. **Eduardo C. Arellano:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition. **Caitlin G. McCormack:** Writing – review & editing, Project administration, Investigation, Formal analysis, Data curation. **Vinina Ferreira:** Writing – review & editing, Methodology, Investigation. **Jonathan Hillier:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Richard Heathcote:** Writing – review & editing, Software, Methodology, Investigation, Conceptualization. **Henk Kloen:** Writing – review & editing, Methodology, Conceptualization. **Andrés Muñoz-Sáez:** Writing – review & editing, Methodology, Investigation. **Patricia Oliveira-Rebouças:** Methodology, Investigation. **Fabiana Oliveira da Silva:** Writing – review & editing, Methodology, Investigation. **Nadia Rojas-Arévalo:** Writing – review & editing, Methodology, Investigation. **Natalia Zielonka:** Writing – review & editing, Methodology, Investigation. **Lynn V. Dicks:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Lynn Dicks reports a relationship with Cool Farm Alliance that includes: non-financial support. Eduardo Arellano reports a relationship with Cool Farm Alliance that includes: non-financial support and travel reimbursement. The Cool Farm Biodiversity Metric software is managed by Cool Farm Alliance, a community interest company. Registered address: 87b Westgate, Grantham, Lincolnshire, NG31 6LE England. Email: info@coolfarmtool.org. It is co-owned by the Cool Farm Alliance, CLM and the University of Cambridge.

Data availability

The data that has been used is confidential.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2024.111897.

References

[WWW Document], 2022. https://rspo.org/.

Rainforest Alliance [WWW Document], 2022. URL https://www.rainforest-alliance.org/.

Amano, T., González-Varo, J.P., Sutherland, W.J., 2016. Languages are still a major Barrier to global science. PLoS Biol. 14, e2000933.

Batáry, P., Dicks, L.V., Kleijn, D., Sutherland, W.J., 2015. The role of agri-environment schemes in conservation and environmental management. Conserv. Biol. 29, 1006–1016. https://doi.org/10.1111/cobi.12536.

Berthinussen, A., Richardson, O., Altringham, J., 2020. Bat conservation: global synthesis for the effects of interventions. Pegasus Publishing, Exeter, Synopses of Conservation Evidence.

Birrer, S., Zellweger-Fischer, J., Stoeckli, S., Korner-Nievergelt, F., Balmer, O., Jenny, M., Pfiffner, L., 2014. Biodiversity at the farm scale: a novel credit point system. Agr Ecosyst Environ 197, 195–203. https://doi.org/10.1016/j.agee.2014.08.008.

Blaauw, B.R., Isaacs, R., 2014. Flower plantings increase wild bee abundance and the pollination services provided to a pollination-dependent crop. J. Appl. Ecol. 51, 890–898. https://doi.org/10.1111/1365-2664.12257.

Brandt, A.J., MacLeod, C.J., Dicks, L.V., Shackelford, G.E., 2018. Effectiveness of farm management actions for enhancing NZ biodiversity: Specialist judgement assessment The NZ Sustainability Dashboard Research Report 18/06. Published by ARGOS.

Brussaard, L., de Ruiter, P.C., Brown, G.G., 2007. Soil biodiversity for agricultural sustainability. Agr Ecosyst Environ 121, 233–244. https://doi.org/10.1016/j.agee.2006.12.013.

Bull, J.W., Suttle, K.B., Gordon, A., Singh, N.J., Milner-Gulland, E.J., 2013. Biodiversity offsets in theory and practice. Oryx 47, 369–380. https://doi.org/10.1017/S003060531200172X.

Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vie, J.-C., Watson, R., 2010. Global biodiversity: indicators of recent declines. Science 328, 1164–1168. https://doi.org/10.1126/science.1187512.

Clough, Y., Barkmann, J., Juhrbandt, J., Kessler, M., Wanger, T.C., Anshary, A., Buchori, D., Cicuzza, D., Darras, K., Putra, D.D., Erasmi, S., Pitopang, R., Schmidt, C., Schulze, C.H., Seidel, D., Steffan-Dewenter, I., Stenchly, K., Vidal, S., Weist, M., Wielgoss, A.C., Tscharntke, T., 2011. Combining high biodiversity with high yields in tropical agroforests. PNAS 108, 8311–8316. https://doi.org/10.1073/pnas.1016799108.

Clough, Y., Kirchweger, S., Kantelhardt, J., 2020. Field sizes and the future of farmland biodiversity in european landscapes. Conserv. Lett. 13, e12752.

Connor, E.F., McCoy, E.D., 1979. The statistics and biology of the species-area relationship. Am. Nat. 113, 791–833. https://doi.org/10.1086/283438.

Conservation Evidence [WWW Document], 2022. URL https://www.conservationevidence.com/.

Convention on Biological Diversity, 2022. Kunming-Montreal Global Biodiversity Framework (No. CBD/COP/DEC/15/4).

Cook, C.N., Hockings, M., Carter, R., (Bill),, 2010. Conservation in the dark? the information used to support management decisions. Front. Ecol. Environ. 8, 181–186. https://doi.org/10.1890/090020.

Cunningham, S.A., Attwood, S.J., Bawa, K.S., Benton, T.G., Broadhurst, L.M., Didham, R.K., McIntyre, S., Perfecto, I., Samways, M.J., Tscharntke, T., Vandermeer, J., Villard, M.-A., Young, A.G., Lindenmayer, D.B., 2013. To close the yield-gap while saving biodiversity will require multiple locally relevant strategies. Agr Ecosyst Environ 173, 20–27. https://doi.org/10.1016/j.agee.2013.04.007.

Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., Garibaldi, L.A., Ichii, K., Liu, J., Subramanian, S.M., Midgley, G.F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R.R., Shin, Y.-J., Visseren-Hamakers, I., Willis, K.J., Zayas, C.N., 2019. Pervasive human-driven decline of life on Earth points to the need for transformative change. Science 366, eaax3100. https://doi.org/10.1126/science.aax3100.

Dicks, L., Ashpole, J.E., Danhardt, J., James, K., Jonsson, A., Randall, N., Showler, A., Smith, R.K., Turpie, S., Williams, D., Sutherland, W.J., 2013. Farmland conservation: evidence for the effects of interventions in northern and Western Europe. Pelagic Publishing, Exeter.

Dicks, L.V., Walsh, J.C., Sutherland, W.J., 2014. Organising evidence for environmental management decisions: a '4S' hierarchy. Trends Ecol. Evol. 29, 607–613. https://doi.org/10.1016/j.tree.2014.09.004.

Dicks, L.V., Wright, H.L., Ashpole, J.E., Hutchison, J., McCormack, C.G., Livoreil, B., Zulka, K.P., Sutherland, W.J., 2016. What works in conservation? using expert assessment of summarised evidence to identify practices that enhance natural pest control in agriculture. Biodivers. Conserv. 25, 1383–1399. https://doi.org/10.1007/s10531-016-1133-7.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N.D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E.C., Jones, B., Barber, C.V., Hayes, R., Kormos, C., Martin, V., Crist, E., Sechrest, W., Price, L., Baillie, J.E.M., Weeden, D., Suckling, K., Davis, C., Sizer, N., Moore, R., Thau, D., Birch, T., Potapov, P., Turubanova, S., Tyukavina, A., de Souza, N., Pintea, L., Brito, J.C., Llewellyn, O.A., Miller, A.G., Patzelt, A., Ghazanfar, S.A., Timberlake, J., Klöser, H., Shennan-Farpón, Y., Kindt, R., Lillesø, J.-P.-B., van Breugel, P., Graudal, L., Voge, M., Al-Shammari, K.F., Saleem, M., 2017. An ecoregion-based approach to protecting half the terrestrial realm. Bioscience 67, 534–545. https://doi.org/10.1093/biosci/bix014.

Resolve Ecoregions [WWW Document], 2020. URL https://ecoregions.appspot.com/.

EU Life Initiative, 2021. Biodiversity performance tool [WWW document]. accessed 4.16.21. https://www.biodiversity-performance.eu/.

Fischer, J., Abson, D.J., Bergsten, A., French Collier, N., Dorresteijn, I., Hanspach, J., Hylander, K., Schultner, J., Senbeta, F., 2017. Reframing the food-biodiversity challenge. Trends Ecol. Evol. 32, 335–345. https://doi.org/10.1016/j.tree.2017.02.009.

Franks, J.R., 2019. An assessment of the landscape-scale dimensions of land based environmental management schemes offered to farmers in England. Land Use Policy 83, 147–159. https://doi.org/10.1016/j.landusepol.2019.01.044.

Gabel, V., Home, R., Stöckli, S., Meier, M., Stolze, M., Köpke, U., 2018. Evaluating on-Farm biodiversity: a Comparison of assessment methods. Sustainability 10. https://doi.org/10.3390/su10124812.

Gamez-Virues, S., Perovic, D., Gossner, M., Borsching, C., Bluthgen, N., de Jong, H., Simons, N., Klein, A., Krauss, J., Maier, G., Scherber, C., Steckel, J., Rothenwohrer, C., Steffan-Dewenter, I., Weiner, C., Weisser, W., Werner, M., Tscharntke, T., Westphal, C., 2015. Landscape simplification filters species traits and drives biotic homogenization. Nat. Commun. 6 https://doi.org/10.1038/ncomms9568.

Garibaldi, L.A., Carvalheiro, L.G., Leonhardt, S.D., Aizen, M.A., Blaauw, B.R., Isaacs, R., Kuhlmann, M., Kleijn, D., Klein, A.M., Kremen, C., Morandin, L., Scheper, J., Winfree, R., 2014. From research to action: enhancing crop yield through wild pollinators. Front. Ecol. Environ. 12, 439–447. https://doi.org/10.1890/130330.

Garibaldi, L.A., Oddi, F.J., Miguez, F.E., Bartomeus, I., Orr, M.C., Jobbágy, E.G., Kremen, C., Schulte, L.A., Hughes, A.C., Bagnato, C., Abramson, G., Bridgewater, P., Carella, D.G., Díaz, S., Dicks, L.V., Ellis, E.C., Goldenberg, M., Huaylla, C.A., Kuperman, M., Locke, H., Mehrabi, Z., Santibañez, F., Zhu, C.-D., 2021. Working landscapes need at least 20% native habitat. Conserv. Lett. 14, e12773.

Gerten, D., Heck, V., Jägermeyr, J., Bodirsky, B.L., Fetzer, I., Jalava, M., Kummu, M., Lucht, W., Rockström, J., Schaphoff, S., Schellnhuber, H.J., 2020. Feeding ten billion people is possible within four terrestrial planetary boundaries. Nat. Sustainability 3, 200–208. https://doi.org/10.1038/s41893-019-0465-1.

Giakoumi, S., McGowan, J., Mills, M., Beger, M., Bustamante, R.H., Charles, A., Christie, P., Fox, M., Garcia-Borboroglu, P., Gelcich, S., Guidetti, P., Mackelworth, P., Maina, J.M., McCook, L., Micheli, F., Morgan, L.E., Mumby, P.J., Reyes, L.M., White, A., Grorud-Colvert, K., Possingham, H.P., 2018. Revisiting "success" and "failure" of Marine protected areas: a conservation scientist perspective. Front. Mar. Sci. 5 https://doi.org/10.3389/fmars.2018.00223.

Gibson, L., Lee, T.M., Koh, L.P., Brook, B.W., Gardner, T.A., Barlow, J., Peres, C.A., Bradshaw, C.J.A., Laurance, W.F., Lovejoy, T.E., Sodhi, N.S., 2011. Primary forests are irreplaceable for sustaining tropical biodiversity. Nature 478, 378–381. https://doi.org/10.1038/nature10425.

Global GAP [WWW Document], 2022. Welcome to GLOBALG.A.P. – a Trademark and Set of Standards for Good Agricultural Practices. URL https://www.globalgap.org/.

GloBio [WWW Document], 2022. Global biodiversity model for policy support. URL https://www.globio.info/.

Haddaway, N.R., Kohl, C., Rebelo da Silva, N., Schiemann, J., Spök, A., Stewart, R., Sweet, J.B., Wilhelm, R., 2017. A framework for stakeholder engagement during systematic reviews and maps in environmental management. Environ Evid 6, 11. https://doi.org/10.1186/s13750-017-0089-8.

Hanski, I., 1998. Metapopulation dynamics. Nature 396, 41–49. https://doi.org/10.1038/23876.

Harrison, P.A., Berry, P.M., Simpson, G., Haslett, J.R., Blicharska, M., Bucur, M., Dunford, R., Egoh, B., Garcia-Llorente, M., Geamănă, N., Geertsema, W., Lommelen, E., Meiresonne, L., Turkelboom, F., 2014. Linkages between biodiversity attributes and ecosystem services: a systematic review. Ecosyst. Serv. 9, 191–203. https://doi.org/10.1016/j.ecoser.2014.05.006.

Herd-Hoare, S., Shackleton, C.M., 2020. Ecosystem disservices matter when valuing ecosystem benefits from small-scale arable agriculture. Ecosyst. Serv. 46, 101201 https://doi.org/10.1016/j.ecoser.2020.101201.

Hodge, I., Reader, M., 2010. The introduction of entry level Stewardship in England: extension or dilution in agri-environment policy? Land Use Policy, Forest Transitions 27, 270–282. https://doi.org/10.1016/j.landusepol.2009.03.005.

IUCN Red List [WWW Document], 2022. URL https://www.iucnredlist.org/.

Jolibert, C., Wesselink, A., 2012. Research impacts and impact on research in biodiversity conservation: the influence of stakeholder engagement. Environ Sci Policy 22, 100–111. https://doi.org/10.1016/j.envsci.2012.06.012.

Kadykalo, A.N., Buxton, R.T., Morrison, P., Anderson, C.M., Bickerton, H., Francis, C.M., Smith, A.C., Fahrig, L., 2021. Bridging research and practice in conservation. Conservation Biology N/a. https://doi.org/10.1111/cobi.13732.

Kainer, K.A., DiGiano, M.L., Duchelle, A.E., Wadt, L.H.O., Bruna, E., Dain, J.L., 2009. Partnering for greater success: local stakeholders and Research in tropical biology and conservation. Biotropica 41, 555–562. https://doi.org/10.1111/j.1744-7429.2009.00560.x.

Karp, D.S., Chaplin-Kramer, R., Meehan, T.D., Martin, E.A., DeClerck, F., Grab, H., Gratton, C., Hunt, L., Larsen, A.E., Martínez-Salinas, A., O'Rourke, M.E., Rusch, A., Poveda, K., Jonsson, M., Rosenheim, J.A., Schellhorn, N.A., Tscharntke, T., Wratten, S.D., Zhang, W., Iverson, A.L., Adler, L.S., Albrecht, M., Alignier, A., Angelella, G.M., Zubair Anjum, M., Avelino, J., Batáry, P., Baveco, J.M., Bianchi, F.J.J.A., Birkhofer, K., Bohnenblust, E.W., Bommarco, R., Brewer, M.J., Caballero-López, B., Carrière, Y., Carvalheiro, L.G., Cayuela, L., Centrella, M., Ćetković, A., Henri, D.C., Chabert, A., Costamagna, A.C., De la Mora, A., de Kraker, J., Desneux, N., Diehl, E., Diekötter, T., Dormann, C.F., Eckberg, J.O., Entling, M.H., Fiedler, D., Franck, P., Frank van Veen, F.J., Frank, T., Gagic, V., Garratt, M.P.D., Getachew, A., Gonthier, D.J., Goodell, P.B., Graziosi, I., Groves, R.L., Gurr, G.M., Hajian-Forooshani, Z., Heimpel, G.E., Herrmann, J.D., Huseth, A.S., Inclán, D.J., Ingrao, A.J., Iv, P., Jacot, K., Johnson, G.A., Jones, L., Kaiser, M., Kaser, J.M., Keasar, T., Kim, T.N., Kishinevsky, M., Landis, D.A., Lavandero, B., Lavigne, C., Le Ralec, A., Lemessa, D., Letourneau, D.K., Liere, H., Lu, Y., Lubin, Y., Luttermoser, T., Maas, B., Mace, K., Madeira, F., Mader, V., Cortesero, A.M., Marini, L., Martinez, E., Martinson, H.M., Menozzi, P., Mitchell, M.G.E., Miyashita, T., Molina, G.A.R., Molina-Montenegro, M.A., O'Neal, M.E., Opatovsky, I., Ortiz-Martinez, S., Nash, M., Östman, Ö., Ouin, A., Pak, D., Paredes, D., Parsa, S., Parry, H., Perez-Alvarez, R., Perović, D.J., Peterson, J.A., Petit, S., Philpott, S.M., Plantegenest, M., Plećaš, M., Pluess, T., Pons, X., Potts, S.G., Pywell, R.F., Ragsdale, D.W., Rand, T.A., Raymond, L., Ricci, B., Sargent, C., Sarthou, J.-P., Saulais, J., Schäckermann, J., Schmidt, N.P., Schneider, G., Schüepp, C., Sivakoff, F.S., Smith, H.G., Stack Whitney, K., Stutz, S., Szendrei, Z., Takada, M.B., Taki, H., Tamburini, G., Thomson, L.J., Tricault, Y., Tsafack, N., Tschumi, M., Valantin-Morison, M., Van Trinh, M., van der Werf, W., Vierling, K.T., Werling, B.P., Wickens, J.B., Wickens, V.J., Woodcock, B.A., Wyckhuys, K., Xiao, H., Yasuda, M., Yoshioka, A., Zou, Y., 2018. Crop pests and predators exhibit inconsistent responses to surrounding landscape composition. Proc Natl Acad Sci USA 115, E7863. https://doi.org/10.1073/pnas.1800042115.

Kennedy, C.M., Lonsdorf, E., Neel, M.C., Williams, N.M., Ricketts, T.H., Winfree, R., Bommarco, R., Brittain, C., Burley, A.L., Cariveau, D., Carvalheiro, L.G., Chacoff, N.P., Cunningham, S.A., Danforth, B.N., Dudenhöffer, J.-H., Elle, E., Gaines, H.R., Garibaldi, L.A., Gratton, C., Holzschuh, A., Isaacs, R., Javorek, S.K., Jha, S., Klein, A.M., Krewenka, K., Mandelik, Y., Mayfield, M.M., Morandin, L., Neame, L.A., Otieno, M., Park, M., Potts, S.G., Rundlöf, M., Saez, A., Steffan-Dewenter, I., Taki, H., Viana, B.F., Westphal, C., Wilson, J.K., Greenleaf, S.S., Kremen, C., 2013. A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. Ecol. Lett. 16, 584–599. https://doi.org/10.1111/ele.12082.

Key, G., Whitfield, M., Dicks, L., Sutherland, W.J., Bardgett, R.D., 2013. Enhancing soil fertility as an ecosystem Service: evidence for the effects of actions. The University of Cambridge, Cambridge.

Kloen, H., 2014. Gaia Biodiversity Yardstick [WWW Document]. URL https://gaia-biodiversity-yardstick.eu/ (accessed 4.12.21).

Konvicka, M., Benes, J., Cizek, O., Kopecek, F., Konvicka, O., Vitaz, L., 2008. How too much care kills species: grassland reserves, agri-environmental schemes and extinction of Colias myrmidone (Lepidoptera: pieridae) from its former stronghold. J. Insect Conserv. 12, 519–525. https://doi.org/10.1007/s10841-007-9092-7.

Kremen, C., Merenlender, A.M., 2018. Landscapes that work for biodiversity and people. Science 362. https://doi.org/10.1126/science.aau6020.

Lambertini, M., 2020. Living Planet Report 2020: Bending the Curve of Biodiversity Loss.

Leclère, D., Obersteiner, M., Barrett, M., Butchart, S.H.M., Chaudhary, A., De Palma, A., DeClerck, F.A.J., Di Marco, M., Doelman, J.C., Dürauer, M., Freeman, R., Harfoot, M., Hasegawa, T., Hellweg, S., Hilbers, J.P., Hill, S.L.L., Humpenöder, F., Jennings, N., Krisztin, T., Mace, G.M., Ohashi, H., Popp, A., Purvis, A., Schipper, A.M., Tabeau, A., Valin, H., van Meijl, H., van Zeist, W.-J., Visconti, P., Alkemade, R., Almond, R., Bunting, G., Burgess, N.D., Cornell, S.E., Di Fulvio, F., Ferrier, S., Fritz, S., Fujimori, S., Grooten, M., Harwood, J., Havlík, P., Herrero, M., Hoskins, A.J., Jung, M., Kram, T., Lotze-Campen, H., Matsui, T., Meyer, C., Nel, D., Newbold, T., Schmidt-Traub, G., Stehfest, E., Strassburg, B.B.N., van Vuuren, D.P., Ware, C., Watson, J.E.M., Wu, W., Young, L., 2020. Bending the curve of terrestrial biodiversity needs an integrated strategy. Nature. https://doi.org/10.1038/s41586-020-2705-y.

Li, C., Hoffland, E., Kuyper, T.W., Yu, Y., Zhang, C., Li, H., Zhang, F., van der Werf, W., 2020. Syndromes of production in intercropping impact yield gains. Nat. Plants 6, 653–660. https://doi.org/10.1038/s41477-020-0680-9.

Lynch, A.J., Fernández-Llamazares, Á., Palomo, I., Jaureguiberry, P., Amano, T., Basher, Z., Lim, M., Mwampamba, T.H., Samakov, A., Selomane, O., 2021. Culturally diverse expert teams have yet to bring comprehensive linguistic diversity to intergovernmental ecosystem assessments. One Earth 4, 269–278. https://doi.org/10.1016/j.oneear.2021.01.002.

MacLeod, C.J., Brandt, A.J., Collins, K., Dicks, L.V., 2022a. Giving stakeholders a voice in governance: biodiversity priorities for New Zealand's agriculture. People and Nature 4, 330–350. https://doi.org/10.1002/pan3.10285.

MacLeod, C.J., Brandt, A.J., Dicks, L.V., 2022b. Facilitating the wise use of experts and evidence to inform local environmental decisions. People and Nature 4, 904–917. https://doi.org/10.1002/pan3.10328.

MacLeod, C. J., Brandt, A. J., Collins, K., Dicks, L. V., Green, P., & Shackelford, G. E., 2018. Biodiversity assessment tool. Prototype version 1.0. Shiny application. 25 May 2018.

Martin, E.A., Dainese, M., Clough, Y., Báldi, A., Bommarco, R., Gagic, V., Garratt, M.P.D., Holzschuh, A., Kleijn, D., Kovács-Hostyánszki, A., Marini, L., Potts, S.G., Smith, H.G., Al Hassan, D., Albrecht, M., Andersson, G.K.S., Asís, J.D., Aviron, S., Balzan, M.V., Baños-Picón, L., Bartomeus, I., Batáry, P., Burel, F., Caballero-López, B., Concepción, E.D., Coudrain, V., Dänhardt, J., Diaz, M., Diekötter, T., Dormann, C.F., Duflot, R., Entling, M.H., Farwig, N., Fischer, C., Frank, T., Garibaldi, L.A., Hermann, J., Herzog, F., Inclán, D., Jacot, K., Jauker, F., Jeanneret, P., Kaiser, M., Krauss, J., Le Féon, V., Marshall, J., Moonen, A., Moreno, G., Riedinger, V., Rundlöf, M., Rusch, A., Scheper, J., Schneider, G., Schüepp, C., Stutz, S., Sutter, L., Tamburini, G., Thies, C., Tormos, J., Tscharntke, T., Tschumi, M., Uzman, D., Wagner, C., Zubair-Anjum, M., Steffan-Dewenter, I., 2019. The interplay of landscape composition and configuration: new pathways to manage functional biodiversity and agroecosystem services across Europe. Ecol Lett 22, 1083–1094. https://doi.org/10.1111/ele.13265.

Meixler, M.S., Fisher, K., Sanderson, E.W., 2019. Latitude-enhanced species-area relationships for conservation planning. Landscape Ecol 34, 1877–1888. https://doi. org/10.1007/s10980-019-00863-2.

NATCAP MAP [WWW Document], 2022. URL https://www.natcapresearch.com/latest-tools.

Nuñez, M.A., Amano, T., 2021. Monolingual searches can limit and bias results in global literature reviews. Nat. Ecol. Evol. 5, 264. https://doi.org/10.1038/s41559-020-01369-w.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R., 2001. Terrestrial ecoregions of the world: a new map of life on Earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. Bioscience 51, 933–938. https://doi.org/10.1641/0006-3568(2001) 051[0933:TEOTWA]2.0.CO;2.

Pashkevich, M.D., d'Albertas, F., Aryawan, A.A.K., Buchori, D., Caliman, J.-P., Chaves, A. D.G., Hidayat, P., Kreft, H., Naim, M., Razafimahatratra, A., Turner, E.C., Zemp, D. C., Luke, S.H., 2022. Nine actions to successfully restore tropical agroecosystems. Trends Ecol. Evol. https://doi.org/10.1016/j.tree.2022.07.007.

Phalan, B.T., 2018. What have we Learned from the land Sparing-sharing model? Sustainability 10. https://doi.org/10.3390/su10061760.

Putman, R., Dunn, N., Zhang, E., Chen, M., Miquel, C., Savolainen, V., 2021. Conservation genetics of native and european-introduced chinese water deer (Hydropotes inermis). Zool. J. Linn. Soc. 191, 1181–1191. https://doi.org/10.1093/ zoolinnean/zlaa076.

Pywell, R.F., Heard, M.S., Woodcock, B.A., Hinsley, S., Ridding, L., Nowakowski, M., Bullock, J.M., 2015. Wildlife-friendly farming increases crop yield: evidence for ecological intensification. Proc. R. Soc. B Biol. Sci. 282, 20151740. https://doi.org/ 10.1098/rspb.2015.1740.

Redpath, S.M., Young, J., Evely, A., Adams, W.M., Sutherland, W.J., Whitehouse, A., Amar, A., Lambert, R.A., Linnell, J.D.C., Watt, A., Gutiérrez, R.J., 2013. Understanding and managing conservation conflicts. Trends Ecol. Evol. 28, 100–109. https://doi.org/10.1016/j.tree.2012.08.021.

Reed, M.S., Vella, S., Challies, E., de Vente, J., Frewer, L., Hohenwallner-Ries, D., Huber, T., Neumann, R.K., Oughton, E.A., Sidoli del Ceno, J., van Delden, H., 2018. A theory of participation: what makes stakeholder and public engagement in environmental management work? Restor. Ecol. 26, S7–S17. https://doi.org/ 10.1111/rec.12541.

Rigal, S., Dakos, V., Alonso, H., Auniņš, A., Benkő, Z., Brotons, L., Chodkiewicz, T., Chylarecki, P., de Carli, E., del Moral, J.C., Domşa, C., Escandell, V., Fontaine, B., Foppen, R., Gregory, R., Harris, S., Herrando, S., Husby, M., Ieronymidou, C., Jiguet, F., Kennedy, J., Klvaňová, A., Kmecl, P., Kuczyński, L., Kurlavičius, P., Kålås, J.A., Lehikoinen, A., Lindström, Å., Lorrillière, R., Moshøj, C., Nellis, R., Noble, D., Eskildsen, D.P., Paquet, J.-Y., Pélissié, M., Pladevall, C., Portolou, D., Reif, J., Schmid, H., Seaman, B., Szabo, Z.D., Szép, T., Florenzano, G.T., Teufelbauer, N., Trautmann, S., van Turnhout, C., Vermouzek, Z., Vikstrøm, T., Voříšek, P., Weiserbs, A., Devictor, V., 2023. Farmland practices are driving bird population decline across Europe. Proceedings of the National Academy of Sciences 120, e2216573120. 10.1073/pnas.2216573120.

Riva, F., Fahrig, L., 2023. Landscape-scale habitat fragmentation is positively related to biodiversity, despite patch-scale ecosystem decay. Ecol. Lett. 26, 268–277. https:// doi.org/10.1111/ele.14145.

Rose, D.C., Sutherland, W.J., Parker, C., Lobley, M., Winter, M., Morris, C., Twining, S., Ffoulkes, C., Amano, T., Dicks, L.V., 2016. Decision support tools for agriculture: Towards effective design and delivery. Agr. Syst. 149, 165–174. https://doi.org/ 10.1016/j.agsy.2016.09.009.

Rose, D.C., Parker, C., Fodey, J., Park, C., Sutherland, W.J., Dicks, L.V., 2017. Involving stakeholders in agricultural decision support systems: improving user-centred design. International Journal of Agricultural Management 6, 80–89. https://doi.org/ 10.5836/ijam/2017-06-80.

Schader, C., Drapela, T., Markut, T., Meier, M.S., Lindenthal, T., Hörtenhuber, S., Pfiffner, L., 2014. Farm- and Product-Level Biodiversity Assessment of Conventional and Organic Dairy Production in Austria. null 10, 20–39. https://doi.org/10.1080/ 21513732.2013.878752.

Schader, C., Baumgart, L., Landert, J., Muller, A., Ssebunya, B., Blockeel, J., Weisshaidinger, R., Petrasek, R., Mészáros, D., Padel, S., Gerrard, C., Smith, L., Lindenthal, T., Niggli, U., Stolze, M., 2016. Using the sustainability monitoring and assessment routine (SMART) for the systematic analysis of trade-offs and synergies

between sustainability dimensions and themes at Farm level. Sustainability 8. https://doi.org/10.3390/su8030274.

Scheper, J., Holzschuh, A., Kuussaari, M., Potts, S., Rundlof, M., Smith, H., Kleijn, D., 2013. Environmental factors driving the effectiveness of european agri-environmental measures in mitigating pollinator loss - a meta-analysis. Ecol. Lett. 16, 912–920. https://doi.org/10.1111/ele.12128.

Schipper, A.M., Hilbers, J.P., Meijer, J.R., Antão, L.H., Benítez-López, A., de Jonge, M.M. J., Leemans, L.H., Scheper, E., Alkemade, R., Doelman, J.C., Mylius, S., Stehfest, E., van Vuuren, D.P., van Zeist, W.-J., Huijbregts, M.A.J., 2020. Projecting terrestrial biodiversity intactness with GLOBIO 4. Glob. Chang. Biol. 26, 760–771. https://doi. org/10.1111/gcb.14848.

Schneider, M.K., Lüscher, G., Jeanneret, P., Arndorfer, M., Ammari, Y., Bailey, D., Balázs, K., Báldi, A., Choisis, J.-P., Dennis, P., Eiter, S., Fjellstad, W., Fraser, M.D., Frank, T., Friedel, J.K., Garchi, S., Geijzendorffer, I.R., Gomiero, T., Gonzalez-Bornay, G., Hector, A., Jerkovich, G., Jongman, R.H.G., Kakudidi, E., Kainz, M., Kovács-Hostyánszki, A., Moreno, G., Nkwiine, C., Opio, J., Oschatz, M.-L., Paoletti, M.G., Pointereau, P., Pulido, F.J., Sarthou, J.-P., Siebrecht, N., Sommaggio, D., Turnbull, L.A., Wolfrum, S., Herzog, F., 2014. Gains to species diversity in organically farmed fields are not propagated at the farm level. Nat. Commun. 5, 4151. https://doi.org/10.1038/ncomms5151.

Science Based Targets Network [WWW Document], 2022. URL https://www.nature. com/articles/s41477-020-0680-9 (accessed 5.8.21).

Searchinger, T., Waite, R., Hanson, C., Ranganathan, J., Dumas, P., Matthews, E., 2019. Creating a Sustainable Future: A Menu of Solutions to Feed Nearly 10 Billion People by 2050.

Seibold, S., Gossner, M., Simons, N., Bluthgen, N., Muller, J., Ambarli, D., Ammer, C., Bauhus, J., Fischer, M., Habel, J., Linsenmair, K., Nauss, T., Penone, C., Prati, D., Schall, P., Schulze, E., Vogt, J., Wollauer, S., Weisser, W., 2019. Arthropod decline in grasslands and forests is associated with landscape-level drivers. Nature 574, 671-+. https://doi.org/10.1038/s41586-019-1684-3.

Shackelford, G.E., Kelsey, R., Robertson, R.J., Williams, D.R., Dicks, L.V., 2017. Sustainable agriculture in California and Mediterranean climates: evidence for the effects of selected interventions. University of Cambridge, Cambridge, Synopses of Conservation Evidence Series.

Shackelford, G.E., Kelsey, R., Sutherland, W.J., Kennedy, C.M., Wood, S.A., Gennet, S., Karp, D.S., Kremen, C., Seavy, N.E., Jedlicka, J.A., Gravuer, K., Kross, S.M., Bossio, D.A., Muñoz-Sáez, A., LaHue, D.G., Garbach, K., Ford, L.D., Felice, M., Reynolds, M.D., Rao, D.R., Boomer, K., LeBuhn, G., Dicks, L.V., 2019. Evidence synthesis as the basis for Decision analysis: a method of selecting the best agricultural Practices for multiple ecosystem Services. Front. Sustain. Food Syst. 3 https://doi.org/10.3389/fsufs.2019.00083.

Shackelford, G.E., Martin, P.A., Hood, A.S.C., Christie, A.P., Kulinskaya, E., Sutherland, W.J., 2021. Dynamic meta-analysis: a method of using global evidence for local decision making. BMC Biol. 19, 33. https://doi.org/10.1186/s12915-021-00974-w.

Sietz, D., Klimek, S., Dauber, J., 2022. Tailored pathways toward revived farmland biodiversity can inspire agroecological action and policy to transform agriculture. Communications Earth & Environment 3, 211. https://doi.org/10.1038/s43247-022-00527-1.

Simmonds, J.S., Sonter, L.J., Watson, J.E.M., Bennun, L., Costa, H.M., Dutson, G., Edwards, S., Grantham, H., Griffiths, V.F., Jones, J.P.G., Kiesecker, J., Possingham, H.P., Puydarrieux, P., Quétier, F., Rainer, H., Rainey, H., Roe, D., Savy, C.E., Souquet, A., ten Kate, K., Victurine, R., von Hase, A., Maron, M., 2020. Moving from biodiversity offsets to a target-based approach for ecological compensation. Conserv. Lett. 13, e12695.

Smith, D.G., 2017. Unilever Sustainable Agriculture Code 2015. Sustainable Agriculture 36.

Sterling, E.J., Betley, E., Sigouin, A., Gomez, A., Toomey, A., Cullman, G., Malone, C., Pekor, A., Arengo, F., Blair, M., Filardi, C., Landrigan, K., Porzecanski, A.L., 2017. Assessing the evidence for stakeholder engagement in biodiversity conservation. Biol. Conserv. 209, 159–171. https://doi.org/10.1016/j.biocon.2017.02.008.

Sustainable Agriculture Initiative, 2021. Sustainable agriculture principles and Practices [WWW document]. accessed 4.15.21 SAI Platform. https://saiplatform.org/our-work/reports-publications/sustainable-agriculture-principles-and-practices/.

Sustainable Agriculture Initiative, 2014. Sustainability Performance Assessment Version 2.0. Towards Consistent Measurement of Sustainability at Farm Level, Sustainable Agriculture Initiative (SAI) Platform.

Sutherland, W.J., Pullin, A.S., Dolman, P.M., Knight, T.M., 2004. The need for evidence-based conservation. Trends Ecol. Evol. 19, 305–308. https://doi.org/10.1016/j. tree.2004.03.018.

Sutherland, W.J., Taylor, N.G., MacFarlane, D., Amano, T., Christie, A.P., Dicks, L.V., Lemasson, A.J., Littlewood, N.A., Martin, P.A., Ockendon, N., Petrovan, S.O., Robertson, R.J., Rocha, R., Shackelford, G.E., Smith, R.K., Tyler, E.H.M., Wordley, C. F.R., 2019. Building a tool to overcome barriers in research-implementation spaces: the conservation evidence database. Biol. Conserv. 238, 108199 https://doi.org/ 10.1016/j.biocon.2019.108199.

Sutherland, W.J., Wordley, C.F.R., 2017. Evidence complacency hampers conservation. Nat. Ecol. Evol. 1, 1215–1216. https://doi.org/10.1038/s41559-017-0244-1.

Tamburini, G., Bommarco, R., Wanger, T.C., Kremen, C., van der Heijden, M.G.A., Liebman, M., Hallin, S., 2020. Agricultural diversification promotes multiple ecosystem services without compromising yield. science. Advances 6, eaba1715. https://doi.org/10.1126/sciadv.aba1715.

Tasser, E., Rüdisser, J., Plaikner, M., Wezel, A., Stöckli, S., Vincent, A., Nitsch, H., Dubbert, M., Moos, V., Walde, J., Bogner, D., 2019. A simple biodiversity assessment scheme supporting nature-friendly farm management. Ecol. Ind. 107, 105649 https://doi.org/10.1016/j.ecolind.2019.105649.

The United Nations General Assembly, 2015. Transforming our world: the 2030 agenda for sustainable development [WWW document]. accessed 9.10.22. https://sdgs.un.org/2030agenda.

Tilman, D., Hill, J., Lehman, C., 2006. Carbon-negative biofuels from low-input high-diversity grassland biomass. Science 314, 1598–1600. https://doi.org/10.1126/science.1133306.

Tscharntke, T., Tylianakis, J.M., Rand, T.A., Didham, R.K., Fahrig, L., Batáry, P., Bengtsson, J., Clough, Y., Crist, T.O., Dormann, C.F., Ewers, R.M., Fründ, J., Holt, R. D., Holzschuh, A., Klein, A.M., Kleijn, D., Kremen, C., Landis, D.A., Laurance, W., Lindenmayer, D., Scherber, C., Sodhi, N., Steffan-Dewenter, I., Thies, C., van der Putten, W.H., Westphal, C., 2012. Landscape moderation of biodiversity patterns and processes - eight hypotheses. Biol. Rev. 87, 661–685. https://doi.org/10.1111/j.1469-185X.2011.00216.x.

Tuck, S.L., Winqvist, C., Mota, F., Ahnström, J., Turnbull, L.A., Bengtsson, J., 2014. Land-use intensity and the effects of organic farming on biodiversity: a hierarchical meta-analysis. J. Appl. Ecol. 51, 746–755. https://doi.org/10.1111/1365-2664.12219.

Tzilivakis, J., Warner, D.J., Green, A., Lewis, K.A., Angileri, V., 2016. An indicator framework to help maximise potential benefits for ecosystem services and biodiversity from ecological focus areas. Ecol. Ind. 69, 859–872. https://doi.org/10.1016/j.ecolind.2016.04.045.

van Doorn, A., Jongeneel, R., 2020. Exploring the possibilities for implementation of the dutch biodiversity monitor for dairy and arable farming in the CAP after 2022. Wageningen, Wageningen Environmental Research.

Watson, J.E.M., Dudley, N., Segan, D.B., Hockings, M., 2014. The performance and potential of protected areas. Nature 515, 67–73. https://doi.org/10.1038/nature13947.

Wezel, A., Vincent, A., Nitsch, H., Schmid, O., Dubbert, M., Tasser, E., Fleury, P., Stöckli, S., Stolze, M., Bogner, D., 2018. Farmers' perceptions, preferences, and propositions for result-oriented measures in mountain farming. Land Use Policy 70, 117–127. https://doi.org/10.1016/j.landusepol.2017.10.020.

Williams, D.R., Pople, R.G., Showler, D.A., Dicks, L.V., Child, M.F., Sutherland, W.J., 2012. Bird conservation: global evidence for the effects of interventions. Pelagic Publishing, Exeter, Synopses of Conservation Evidence.

Wright, H.L., Ashpole, J.E., Dicks, L.V., Hutchison, J., Sutherland, W.J., 2013. Enhancing natural Pest control as an ecosystem Service: evidence for the effects of selected actions. Univeristy of Cambridge, Cambridge.

WWF Ecoregions [WWW Document], 2022. URL https://www.worldwildlife.org/biomes.

Young, K.D., Van Aarde, R.J., 2011. Science and elephant management decisions in South Africa. Biol. Conserv. 144, 876–885. https://doi.org/10.1016/j.biocon.2010.11.023.

Zellweger-Fischer, J., Althaus, P., Birrer, S., Jenny, M., Pfiffner, L., Stockli, S., 2016. Survey of biodiversity on Farm business with a point-based system. AGRARFORSCHUNG SCHWEIZ 7, 40–47.